

# Aplicações de *Big Data* e algoritmos de *Machine Learning* à gestão inteligente da rega

**Pedro Dinis Batista Azevedo**

Dissertação para a obtenção do grau de Mestre em  
**Engenharia Agronómica- Especialização em Hortofruticultura e  
Viticultura**

Orientador: Maria Isabel Freire Ribeiro Ferreira

**Júri:**

Presidente: Doutora Maria do Rosário da Conceição Cameira, Professora Associada do Instituto Superior de Agronomia da Universidade de Lisboa

Vogais: Doutora Maria Isabel Freire Ribeiro Ferreira, Professora Catedrática do Instituto Superior de Agronomia da Universidade de Lisboa;

Doutor João Rui Rolim Fernandes Machado Lopes, Professor Auxiliar Convidado do Instituto Superior de Agronomia da Universidade de Lisboa



## **AGRADECIMENTOS**

Um profundo agradecimento à Prof.<sup>a</sup> Dra. Maria Isabel Ferreira por toda a paciência, ajuda e transmissão de conhecimento ao longo da elaboração deste trabalho e na preparação do mesmo.

A todos os colegas e parceiros de trabalho com quem tive oportunidade de privar e conhecer durante o período de aulas no ISA, pela vossa disponibilidade e ajuda a alguém que vinha de um meio completamente diferente.

A todos os professores que com o seu conhecimento me marcaram vincadamente o gosto pela agricultura e a vontade de aprender cada vez mais.

À Mariana Dias, pelo apoio em todos os projectos e ambições de vida, por me acompanhar em todos eles e por ter a paciência de corrigir os meus erros ortográficos. Mas acima de tudo pelo amor com o qual tudo fica mais fácil.

A toda a equipa da Hidrosoph com quem partilho os meus dias de trabalho e continuo a aprender todos os dias a ser melhor profissional, pelo vosso apoio moral e força para concluir esta etapa.

Finalmente aos meus pais e avós, por me apoiarem a voltar aos estudos naquilo que sempre gostei. Especialmente ao meu pai e avô por se terem aventurado no mundo da fruticultura e me terem permitido crescer rodeado de árvores de fruto, despoletando em mim uma enorme vontade de seguir nos seus passos.

## RESUMO

O crescimento demográfico e as alterações climáticas são os dois grandes desafios à produção agrícola do século XXI. A pressão de produzir mais com menos recursos implica mudanças nos métodos de produção e na gestão eficiente de cada um.

A utilização de água para rega é um dos principais motores do crescimento vegetal e a sua disponibilidade deve ser assegurada para o futuro. Para tal é necessário reduzir o desperdício e garantir que a rega aplicada é adequada à necessidade das plantas ao longo do ciclo cultural.

A evolução das tecnologias de comunicação e recolha de dados no campo (sensores de humidade do solo, dendrómetros, etc) permitem criar uma nova dinâmica entre o agricultor e a informação disponível para a tomada de decisão.

Neste trabalho tirou-se partido do grande volume de dados disponível numa parcela de olival superintensivo, monitorizada por uma rede sensorial onde se incluía um dendrómetro e uma sonda de humidade do solo. A partir dos dados meteorológicos e sensoriais, recolhidos ao longo de dois anos, procurou-se encontrar uma relação entre os índices derivados da dendrometria e o estado hídrico da planta, utilizando posteriormente esta relação para construir dois algoritmos, uma rede neuronal (ANN) e uma floresta de decisão aleatória (FDA) que consigam prever com base em variáveis de simples obtenção o valor do índice e inferir por essa via o estado hídrico da planta.

A rede neuronal foi depois utilizada como parte de um sistema de aprendizagem reforçada, onde um algoritmo aprendeu a regar autonomamente com base em  $ET_0$ , a evolução do armazenamento de água no solo e os resultados do algoritmo preditivo obtido pela análise da dendrometria.

Os resultados obtidos mostram que as técnicas de “*Big Data*” e ML são adequadas à análise de dados recolhidos no campo, e à criação de ferramentas de apoio à decisão.

**Palavras-chave:** Rega, “*Big Data*”, “*Machine Learning*”, Dendrometria, Estado hídrico da planta

## ABSTRACT

Climate change and the increased rate of population growth are the two main challenges that modern agriculture is facing. The pressure to produce more with less resources requires changes on production methods and their efficient management.

Water for irrigation is one of the most important crop production factors and its availability should be preserved for the future. In order to accomplish this goal, water losses reduction and proper evaluation of crop water requirements is paramount.

Technological advances on communication and data retrieval from the field (soil humidity sensors, dendrometers, etc) allow us to explore a new dynamic between the farmer and the available field information for the decision making process.

This work focus on analysing a big volume of data collected from a sensor network on a superintensive orchard.

A correlation between the plant water status and indexes derived from dendrometry analysis was explored for the data collected over two years. From these results, two machine learning algorithms were built (an artificial neural network and a random forest tree) to predict a dendrometric derived index, based on easily acquirable inputs, allowing for a simple estimation of plant water status.

This estimation was then used as input for a reinforcement learning system, where an agent was taught how to irrigate autonomously based simply on  $ET_0$ , a soil water reserve estimate and the plant water status as predicted by the previously developed algorithm.

The results show that using big data and machine learning techniques on field data result on valuable insights for the farmer and the prospect of building improved decision support systems.

**Keywords:** Irrigation, Big Data, Machine Learning, Dendrometry, Plant Water Status

## ÍNDICE

Agradecimentos.....	I
Resumo .....	II
Abstract .....	III
Índice de figuras .....	VI
Índice de quadros .....	VIII
Índice de equações.....	IX
Listagem de abreviaturas ou siglas.....	X
1 – Introdução.....	1
2 – Introdução Teórica.....	3
2.1 – A agricultura, o seu passado, presente e futuro .....	3
2.1.1 – Os recursos do planeta e o crescimento demográfico .....	3
2.1.2 – A sustentabilidade do mundo agrícola.....	4
2.1.3 - O uso da água na agricultura do século XXI .....	6
2.2 – Estimativa das necessidades hídricas das culturas.....	6
2.2.1 – A evapotranspiração e a equação de Penman-Monteith .....	7
2.2.2 – Coeficiente cultural e coeficiente de stress hídrico .....	8
2.3 – Indicadores do estado hídrico .....	14
2.3.1 – Indicadores do estado hídrico do solo – Sondas .....	14
2.3.2 – Indicadores do estado hídrico da planta – Dendrômetros.....	16
2.4 – IoT, Big Data e agricultura de precisão .....	19
2.4.1 – Um mundo de ligações.....	19
2.4.2 – Dados e o seu potencial para criar valor .....	20
2.4.3 – A agricultura de precisão.....	21
2.5 – <i>Machine Learning</i> , conceitos e principais algoritmos.....	21
2.5.1 – Redes neuronais .....	26
2.5.2 – Aprendizagem reforçada .....	32
2.6 – A oliveira ( <i>Olea europaea</i> L.).....	36
3 – Materiais e Métodos.....	38
3.1 – Características do local.....	38
3.2 – Instrumentos de medição e aquisição de dados.....	39
3.2.1 – Sistemas de aquisição de dados .....	39
3.2.2- Sonda de capacitância Enviropro® (80 cm) .....	40
3.2.3 – Caudalímetro.....	41
3.2.4 – Dendrômetro .....	41
3.2.5 – Estação de medição.....	42
3.3 – Pré-tratamento de dados.....	42
3.4 – Ambiente de desenvolvimento de algoritmos .....	44
3.5 – Metodologia de análise dos dados de dendrometria .....	44

3.5.1 – Floresta de Decisão aleatória .....	47
3.5.2 – ANN .....	48
3.6 – Metodologia de AR para gestão inteligente da rega .....	49
3.6.1 – Ambiente .....	50
3.6.2 – Acções .....	50
3.6.3 – Estados .....	50
3.6.4 – Recompensas .....	51
3.6.5 – Estratégia de Q-Learning .....	52
4 – Resultados.....	53
4.1 – Análise de dendrometria para detecção do stress hídrico .....	53
4.1.1 – 2016 .....	53
4.1.2 – 2017 .....	58
4.2 – Algoritmos ML de previsão do valor de MDS.....	65
4.3 – Gestão inteligente de rega utilizando AR .....	69
5 – Discussão .....	71
6 – Conclusões .....	74
7 – Referências Bibliográficas.....	75

## ÍNDICE DE FIGURAS

Figura 1 – Previsão do crescimento populacional até 2050 (Wallace, 2000).....	4
Figura 2 – Países afectados por seca extrema em 2000 e previsão para 2050 (Gerten et al., 2011). ....	5
Figura 3 - Evolução de Kc ao longo do ciclo cultura (Allen et al., 1998). ....	10
Figura 4 – Processos envolvidos no balanço hídrico do solo. Adaptado de (Allen et al., 1998). ....	11
Figura 5 – Número de dispositivos conectados em rede e projeção até 2025.....	20
Figura 6 – Fluxo de etapas para desenvolvimento de algoritmos de ML.....	23
Figura 7 – Um neurónio cerebral humano (O que é o neurónio?, no date).....	27
Figura 8 – Modelo de um neurónio artificial(Pokharna, 2016). ....	27
Figura 9 – Modelo de uma rede neuronal artificial (How Neural Networks Work, 2017).....	28
Figura 10 – Procura do mínimo absoluto da função de custo(Synced, 2017).....	29
Figura 11 – Treino de uma ANN. ....	30
Figura 12 – Processo de treino detalhado de uma ANN. ....	32
Figura 13 – Modelo de aprendizagem reforçada.....	34
Figura 14 – Ciclo Anual Vegetativo da Oliveira. Adaptado de Tous, 1990 e Rallo, 1998.....	37
Figura 15 - Delimitação da parcela em estudo. ....	38
Figura 16 – Distribuição da precipitação, temperatura máxima e mínima e ET0 entre 2015 e 2018. ....	39
Figura 17 – SAD da estação de medição.....	40
Figura 18 – Sonda Enviropro® antes da instalação. ....	41
Figura 19 – Caudalímetro B-meter.....	41
Figura 20 – Dendrómetro instalado no tronco de uma oliveira da estação de medição. ....	42
Figura 21 – Metodologia de análise da dendrometria.....	45
Figura 22 – Funções de Activação da ANN. ....	48
Figura 23 – Mecanismo de procura do mínimo absoluto da função em 2D. ....	49
Figura 24 – Matriz de Relação das variáveis. ....	54
Figura 25 – Caracterização e Distribuição no tempo de MDS. ....	55
Figura 26 – Caracterização e Distribuição no tempo de TGR. ....	55
Figura 27 –Comparação dos diferentes elementos em estudo ao longo do tempo. De cima para baixo: ASUM, Reserva Teórica, ETa, Ks,MDS e TGR. ....	56
Figura 28 – Comparação de variáveis no período de stress estimado. De cima para baixo: HSUM, MDS e Ks. ....	57
Figura 29 –Distribuição dos valores de MDS pelos períodos em estudo. A amarelo: Não Stress, a azul: Stress Estimado. ....	58
Figura 30 – Matriz de relação das variáveis em estudo. ....	59
Figura 31 –Caracterização e distribuição de MDS no tempo.....	60
Figura 32 – Caracterização e Distribuição de TGR no tempo. ....	60
Figura 33 – Comparação das variáveis em estudo ao longo do tempo. De cima para baixo: HSUM, Reserva Teórica, ETa, Ks, MDS e TGR. ....	61
Figura 34 – Valores no período de Não Stress / Stress. À direita - Período de Stress, À esquerda- Período de Não Stress.....	62
Figura 35 – Valores de HSUM, MDS e Ks no período estimado de stress. ....	62
Figura 36 – Distribuição dos valores de MDS no período de não stress e de stress. A amarelo – Não Stress. A Azul- Stress. ....	63
Figura 37 – Comparação dos anos 2016 e 2017 para as variáveis HSUM, MDS e Ks, nos períodos de stress hídrico estimado. 2016 a azul e 2017 a amarelo. ....	64
Figura 38 – Vista Geral de uma das árvores do algoritmo FDA. ....	66
Figura 39 – Importância das variáveis no resultado final do treino. ....	66
Figura 40 – Dados Reais e Pontos previstos no conjunto de teste.....	67



Figura 41 – Evolução do erro absoluto médio ao longo do processo de treino da ANN .....	67
Figura 42 – Dados reais e pontos previstos pela ANN no conjunto de dados de teste. ....	68
Figura 43 – Histogramas dos erros de previsão de ambos os algoritmos. Esquerda – ANN, Direita – FDA. ....	69
Figura 44 – Recompensa de cada episódio ao longo do processo de treino. Os valores apresentados são a recompensa total, obtida a partir do somatório das recompensas individuais obtidas pela passagem em cada estado durante um episódio.....	70
Figura 45 – Eventos de Rega Simulados e os que efectivamente foram realizados pelo agricultor. Direita- Agricultor, Esquerda - Simulação.....	70
Figura 46 – Variação da reserva de água do solo simulada para o episódio de treino de maior recompensa.....	71

## ÍNDICE DE QUADROS

Quadro 1 – Momentos chave que contribuíram para o desenvolvimento da agricultura (Solheim II, 1972; Faris, 2013).....	3
Quadro 2 – Métodos de determinação da humidade do solo (Vienken et al., 2013)....	15
Quadro 3 – Principais índices derivados da dendrometria (Fernández et al., 2010)....	18
Quadro 4 – Algoritmos comuns de ML (Mitchell, 1997; Shalev-Shwartz et al., 2014; Géron, 2017).....	24
Quadro 5 - Parametrizações de uma Rede Neuronal(Goodfellow et al., 2016; Skansi, 2018) .....	30
Quadro 6 –Especificações da Sonda Enviopro®.....	40
Quadro 7 – Unidades das variáveis recolhidas no SAD. ....	43
Quadro 8 – Valores de Kc ao longo do ciclo cultural. ....	43
Quadro 9 – Bibliotecas e Frameworks utilizadas no trabalho. ....	44
Quadro 10 – Dados recolhidos no campo e estimados do projecto OlivoReg. ....	45
Quadro 11 – Definição dos estados.....	51
Quadro 12 – Valores de recompensa em cada estado .....	51
Quadro 13 – Caracterização estatísticas dos dados utilizados no estudo. ....	53
Quadro 14 – Distribuição dos valores de MDS (mm) nos dois períodos.....	58
Quadro 15 – Caracterização das variáveis recolhidas. ....	58
Quadro 16 – Distribuição de MDS (mm) para os períodos em estudo. ....	63
Quadro 17 – Selecção de um valor base de detecção de Stress Hídrico a partir de MDS(mm).....	63
Quadro 18 – Variáveis de previsão e sua amplitude de valores. ....	65
Quadro 19 – Métricas de Avaliação da FDA. ....	65
Quadro 20 – Parametrizações do sistema de aprendizagem reforçada .....	69

## ÍNDICE DE EQUAÇÕES

1. Equação geral de Penman-Monteith.....	8
2. Equação de Penman- Monteith adaptada para o cálculo da evapotranspiração de referência (FAO).....	8
3. Cálculo da evapotranspiração considerando a presença de stress hídrico na planta .....	11
4. Cálculo da evapotranspiração real considerando $K_s$ como afectando apenas o coeficiente cultural basal(FAO).....	11
5. Cálculo da reserva útil de água no solo.....	12
6. Cálculo da reserva facilmente utilizável.....	12
7. Variação do armazenamento de água no solo para um determinado período de tempo.....	13
8. Condições para a utilização e cálculo do coeficiente de stress.....	13
9. Cálculo do valor de saída de um perceptrão(Goodfellow et al., 2016).....	27
10. Equação das recompensas retardadas.....	34
11. Equação geral de Q-Learning.....	35
12. Factor de correcção para medições obtidas após ajustes no dendrómetro.....	43
13. Conjunto de equações utilizadas para simular a evolução da reserva útil do solo e o valor da evapotranspiração real.....	46
14. Equação de normalização dos valores de uma variável para uma distribuição normal.....	48

## **LISTAGEM DE ABREVIATURAS OU SIGLAS**

IoT – Internet das Coisas

ML – Machine learning

ET – Evapotranspiração

IAF – Índice de Área Foliar

FAO – Food and Agriculture Organization of the United Nations

FAO 56 - FAO Irrigation and drainage paper No. 56

IVDN – Índice de vegetação por diferença normalizada

GD – Graus dia

CC – capacidade de campo

CE – Coeficiente de emurchecimento

TDR – Reflectometria no domínio do tempo

FD – Reflectometria no domínio da frequência

LVDT – Linear Variable Differential Transformer

MDS – Maximum Daily Shrinkage

DR – Daily Recovery

TGR – Trunk Growth Rate

SI – Intensidade de sinal

PC – Computador pessoal

GPS – Global Positioning System

VTA – Variação da taxa de aplicação

ANN – Artificial Neural Network

MIT – Massachusetts Institute of Technology

GPU – Graphics Processing Unit

AR – Aprendizagem reforçada

MDP – Markov Decision Process

GPP – Gabinete de Planeamento, Políticas e Administração Geral

SAD – Sistema de Aquisição de Dados

HTTP – Hypertext Transfer Protocol

FTP – File Transfer Protocol

MQTT – Machine to machine connectivity protocol

SDI-12 – Serial Digital Interface at 1200 baud

BD – Base de dados

RMSE – Root mean square error

FDA – Floresta de Decisão Aleatória

RELU – Rectified Linear Units

AR – Aprendizagem Reforçada

DQN – Deep Q- Network

Por conveniência, neste trabalho usou-se o ponto para separar a parte decimal.

## 1 – INTRODUÇÃO

A agricultura é um dos sectores mais dinâmicos da civilização humana desde que o homem assumiu um estilo de vida sedentário. Várias revoluções agrícolas ao longo dos milénios, baseadas em tecnologias e inovações disruptivas, estiveram na base do crescimento contínuo da população humana, permitindo ao homem subsistir e dedicar-se a outras actividades. Ao longo deste processo evolutivo a agricultura tornou-se cada vez mais intensiva, produzindo uma cada vez maior quantidade de produtos agrícolas na mesma área explorada. Esta metodologia de produção, baseada na aplicação de fitofármacos, fertilizantes sintéticos e rega de conforto, define a agricultura da segunda metade do século XX (Carmo et al., 2017).

No entanto, ocorreram duas grandes mudanças nos finais deste século que põem em causa toda a sustentabilidade da humanidade: o impacto galopante das alterações climáticas e o crescimento exponencial da população.

A consequência principal das alterações climáticas é sentida no aquecimento global e na ocorrência, cada vez mais frequente, de fenómenos extremos, como furacões, tornados e ondas de calor. Todos estes fenómenos tornam as condições climáticas de tal forma instáveis que a viabilidade das culturas fica comprometida. Com a pressão do crescimento demográfico, o desafio colocado ao sector agrícola é produzir mais, com menos recursos e sem esquecer a conservação ambiental. Estas limitações, associadas à diminuição da área arável e à degradação dos solos implicam a necessidade de procurar novos métodos de produção, mais eficientes e onde todo o ciclo cultural possa ser acompanhado, permitindo gerir a cultura em tempo real.

O século XXI trouxe consigo uma profunda revolução tecnológica com a expansão da internet, o aparecimento de redes de comunicação sofisticadas como o 3G e o 4G, ou a criação de processadores cada vez mais pequenos e embutidos em todo o tipo de dispositivos. Estes avanços conduziram ao aparecimento da “*internet of things*” (IoT), um conceito que abrange a existência de um mundo em constante ligação, onde existem redes sensoriais interligadas recolhendo e transmitindo dados, desde as casas inteligentes às parcelas agrícolas. No caso da agricultura, a criação de sensores adaptados a medir variáveis como o teor de humidade do solo ou as variações do diâmetro do tronco abriram a possibilidade de os agricultores terem acesso a informação cada vez mais completa, precisa e acessível a qualquer momento, em qualquer lugar. O tratamento desta informação e a obtenção de conhecimento a partir dela é um tema cada vez mais recorrente entre a comunidade agrícola, sendo, nos países mais desenvolvidos, o valor dos conjuntos de informação designados por “*Big Data*” cada vez maior e existindo até empresas dedicadas à venda de dados adquiridos pelos produtores.

Ao conseguir criar conhecimento sobre as culturas e a sua condução, em ambientes e técnicas culturais completamente distintos, facilitamos a entrada no mercado de novos produtores que assim conseguem consultar informação e adquirir experiência que de outra maneira lhes levaria anos. Para além disso a quantidade de dados gerados diariamente por redes

de IoT permite aos investigadores considerar a utilização de técnicas de “*machine learning*” (ML) e inteligência artificial, cujas potencialidades são já postas em prática noutras áreas.

A nível mundial, um dos principais problemas deste século na prática agrícola é a escassez de água. Como a agricultura é o principal utilizador de água doce, encontra-se sob pressão para reduzir os consumos na gestão das culturas, sendo esta muitas vezes exercida sob a forma de legislação e regulamentos de gestão das bacias. Devido a esta situação, o uso eficiente da água apresenta-se como um dos pontos fulcrais a desenvolver, e sobre o qual recaem muitos dos esforços de investigação.

Se através da informação recolhida nas redes sensoriais instaladas nas culturas conseguirmos extrair conhecimento preciso sobre o estado e as necessidades hídricas dessas culturas em tempo real, será possível quantificar com um nível de incerteza muito mais baixo quais as quantidades de água a aplicar diariamente.

Este trabalho tem como objectivos a análise da viabilidade e precisão de índices derivados a partir de medições de um dendrómetro de tronco na detecção do stress hídrico em oliveiras e também, a construção de uma metodologia de trabalho que permita, a partir de dados recolhidos no campo, construir algoritmos que forneçam indicações do estado hídrico da planta, mesmo em ambiente de simulação computacional. Ao mesmo tempo procura-se a construção de um algoritmo que aprenda por si só a gerir os eventos de rega anuais a partir das necessidades da cultura e dos objectivos de gestão previamente estipulados. Este tipo de algoritmos será facilmente adaptável a outras situações. A criação destes ambientes simulados a partir de dados reais promove a caracterização rápida de cenários que depois são validados no campo, podendo acelerar todo o processo de investigação na área da gestão de rega.

## 2 – INTRODUÇÃO TEÓRICA

### 2.1 – A agricultura, o seu passado, presente e futuro

A agricultura é um sector em constante evolução desde os primórdios das civilizações nómadas. O cultivo de plantas nas zonas mais férteis do planeta e as técnicas de rega rudimentares que as sustentavam levaram ao aparecimento das primeiras comunidades sedentárias, constituindo sem dúvida um dos grandes pilares para o desenvolvimento do ser humano, da sua subsistência e organização em sociedades complexas.

Vários momentos chave na nossa história contribuíram para o desenvolvimento agrícola (Quadro 1):

Quadro 1 – Momentos chave que contribuíram para o desenvolvimento da agricultura (Solheim II, 1972; Faris, 2013).

Ano	Inovação
~10000 BC	Domesticação de plantas e animais (fim das civilizações nómadas)
~1800 AD	Revolução Industrial, mecanização dos processos agrícolas
~1950 AD	"Green Revolution", introdução de OGM's

Ao longo dos séculos, todas estas inovações foram acompanhadas por um aumento da produtividade, novos métodos de combate a pragas e doenças, novas técnicas de rega e novas espécies desenvolvidas pelo homem para sua utilização. Para além disso, estas revoluções na agricultura foram sempre potenciando a nossa evolução enquanto civilização, permitindo ao homem dedicar mais tempo a outras actividades e conhecer novos horizontes enquanto espécie.

Hoje encontramos-nos num ponto de viragem em todas as indústrias, um ponto em que a introdução de máquinas, robôs e técnicas de computação avançadas conseguem superar a performance do ser humano em vários tipos de tarefas e auxiliam técnicos, engenheiros e cientistas a descobrir padrões, comportamentos e modos de operação impensáveis até há bem pouco tempo. A agricultura não é excepção a esta onda de mudança de inovação. Estes avanços tecnológicos são a chave para nos adaptarmos às condições de um planeta em mudança acelerada e profunda.

#### 2.1.1 – Os recursos do planeta e o crescimento demográfico

Desde 1950 a população mundial aumentou 125% (Figura 1), sendo este aumento especialmente acentuado nas regiões asiáticas, particularmente na Índia. Até 2050 é esperado que a população aumente em cerca de 65% (Rijsberman, 2006). Um tão pouco espaço de tempo para um aumento desta magnitude não tem precedentes na história da humanidade e as suas implicações estão ainda a ser debatidas sem a necessária celeridade.

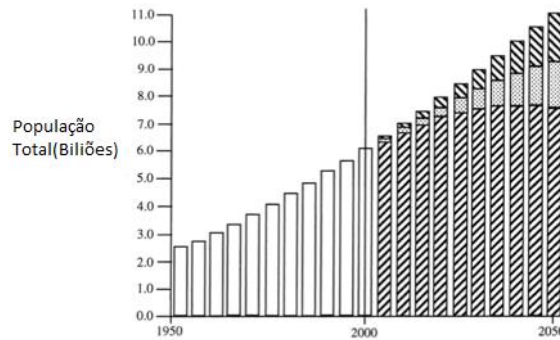


Figura 1 – Previsão do crescimento populacional até 2050 (Wallace, 2000).

Este aumento populacional, aliado às alterações climáticas e aquecimento global apresentam o maior desafio que a indústria agro-alimentar alguma vez enfrentou. Estes fenómenos contribuem para a redução da área arável, também potenciada pelo aumento da poluição, erosão e pela alteração dos ciclos meteorológicos com especial ênfase para a incidência de fenómenos cada vez mais extremos como ondas de calor, inundações, furacões e tempestades (Qadir *et al.*, 2003). A percentagem de água disponível e em condições quer para consumo humano, quer para regadio também vai sofrer um acentuado decréscimo se as previsões actuais se mantiverem. Todos estes processos contribuem para a desregulação e degradação dos ecossistemas o que por sua vez compromete a produção agrícola e a sua sustentabilidade. A curto prazo, e se não forem implementadas novas abordagens à produção vegetal e animal será impossível fornecer as calorias necessárias para a nutrição adequada da maioria da população (Wallace, 2000). Este cenário apresenta-se como uma possibilidade de um futuro sombrio, mas que pode ser atenuado, ou mesmo evitado se tivermos em conta o nível de progresso e engenho que tem marcado as diferentes revoluções agrícolas.

### 2.1.2 – A sustentabilidade do mundo agrícola

Hoje, a agricultura é uma actividade complexa impulsionada pela globalização e pelas necessidades do mercado (procura / oferta). O agricultor, além das ciências agrárias necessita de dominar conhecimentos de áreas tão distintas como as finanças, a biotecnologia e até o marketing. A sua tarefa de procurar maior produtividade e rentabilidade da sua exploração já não é uma relação tão directa como um dia foi. Existem diferentes modos de produção que potenciam ou a qualidade ou a quantidade do produto, sendo, em muitas situações, estas duas variáveis mutuamente exclusivas.

A procura de alimentos aumenta, mas é impossível responder a essa procura com um aumento de áreas cultiváveis. Muitas áreas existentes estão degradadas devido a décadas de agricultura intensiva e aplicação de produtos fitofarmacêuticos sem consideração pelos impactos ambientais a médio/longo prazo (Horrigan *et al.*, 2002; Tilman *et al.*, 2002), ou simplesmente por serem áreas sem condições para a produção agrícola, como acontece em muitas regiões de clima mediterrânico, em que, por condicionantes do clima, especialmente quando associado a



declives elevados, os solos são geralmente pobres e delgados, apresentando-se rapidamente esgotados (Carmo *et al.*, 2017).

Estão a aparecer cada vez mais métodos alternativos de produção, como as culturas hidropónicas sem substrato ou as quintas verticais que ganham popularidade nos grandes centros urbanos, embora sem expressão significativa. Estes métodos e a sua viabilidade só são possíveis devido aos avanços tecnológicos em nutrição das plantas, estratégias de rega, iluminação e design de unidades de produção.

Um dos factores mais limitantes na produção agrícola do futuro é a aplicação eficiente e sustentável da água. Estima-se que 80% dos consumos hídricos se destinem à agricultura (Gerten *et al.*, 2011). Vários estudos têm quantificado a denominada “blue water” que se refere à quantidade de água dos rios, lagos, aquíferos e reservatórios (Gerten *et al.*, 2011). No entanto, a “green water”, a água proveniente da precipitação e que se infiltra no solo e lençóis freáticos é a mais importante no que toca ao crescimento e produtividade dos ecossistemas terrestres, representando a maior parte da água utilizada na agricultura. A escassez de água tende a ser o principal factor limitante à produção (Figura 2).

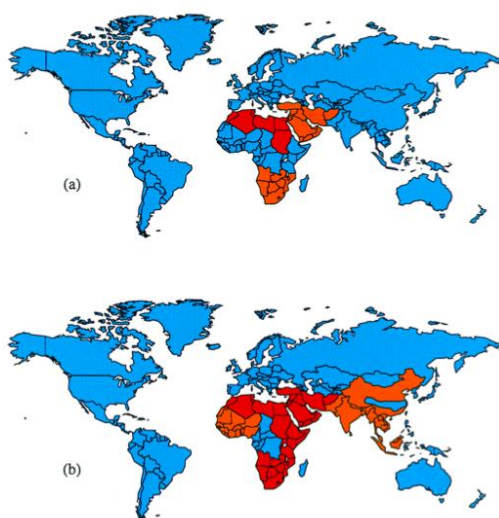


Figura 2 – Países afectados por seca extrema em 2000 e previsão para 2050 (Gerten *et al.*, 2011).

Numa previsão em relação ao nível de água disponível num futuro a médio prazo é possível estimar que a disponibilidade de água na região sul da Europa poderá sofrer reduções de 40% em alguns países, em resposta ao aumento da temperatura e decréscimo da precipitação (Gerten *et al.*, 2011). Mesmo em simulações onde se retira o efeito acelerador da continuada emissão dos gases de estufa é possível inferir que a quantidade de água necessária para produzir uma unidade de biomassa irá aumentar devido ao aquecimento.

Neste momento aproximadamente 7% da população vive em zonas de escassez de água, podendo este número pode aumentar para 67% em 2050 (Molden *et al.*, 2010).

Ou seja, estamos perante um cenário em que é necessário reduzir os consumos devido à escassez de água e ao mesmo tempo melhorar a eficiência na sua aplicação. Precisamos de produzir mais quantidade nas zonas aráveis existentes sem aumentar o consumo de recursos.

Todas estas considerações levam à crescente relevância do uso eficiente e sustentável dos recursos hídricos especialmente aqueles relacionados com a agricultura e que representam a maior fatia dos consumos excessivos.

### **2.1.3 - O uso da água na agricultura do século XXI**

A consciencialização ambiental tem vindo a alterar o panorama da agricultura desde a viragem do século e cada vez mais os agricultores são sensibilizados para o uso eficiente da água, sendo o seu uso cada vez mais regulamentado por parte dos governos e das entidades de gestão.

A agricultura continua a representar a maior fatia dos consumos de água, estimando-se que anualmente represente mais de 70% do consumo total (Vieira, 2003), pelo que a gestão dos recursos hídricos em todo o processo de produção agrícola deve assumir um papel preponderante no planeamento e execução de qualquer projecto de regadio, mesmo tendo em consideração que a maior parte desta água é devolvida à atmosfera por evaporação

Um conceito utilizado muitas vezes neste contexto, é o de eficiência. Torna-se importante então clarificar que eficiência, neste caso, deve ser interpretada como a utilização óptima da água para as necessidades culturais, e não como a poupança água por si só (Pereira *et al.*, 2002). Para melhorar esta eficiência é necessário reduzir a quantidade de água desperdiçada, seja, por exemplo, no armazenamento de água em reservatórios (perdas por evaporação). Cada vez mais esta procura pelas técnicas mais eficientes de regadio é auxiliada pelo número crescente de tecnologias introduzidas em todas as etapas do processo, desde a captação até à quantificação aproximada das necessidades hídricas das plantas.

Esta revolução tecnológica tem vindo a gerar cada vez mais interesse e adopção generalizada com a introdução de tecnologia sensorial no campo, sendo esta cada vez mais fiáveis, o que permite ao agricultor obter e cruzar informação, adquirindo conhecimento em tempo-real, melhorando a sua competitividade. No entanto é necessário tornar este leque de tecnologias acessíveis ao pequeno agricultor, dando-lhe ferramentas simples e completas que lhe forneçam indicações claras de como, quando e quanto regar. Nesse sentido o próximo subcapítulo aborda algumas das técnicas mais utilizadas na estimativa das necessidades hídricas das plantas e da determinação da oportunidade de rega.

## **2.2 – Estimativa das necessidades hídricas das culturas**

A capacidade de estimar as necessidades hídricas das culturas de uma forma precisa através de métodos não invasivos que não perturbem o ecossistema onde as plantas se inserem é um dos grandes avanços na área de gestão eficiente da rega, já ultrapassados há décadas para culturas herbáceas/anuais mas que ainda se mantém um desafio para cobertos de espécies lenhosas (Ferreira *et al.*, 2008). Sem estas metodologias continuaríamos a depender apenas da intuição e experiência acumulada do agricultor ou de equipamentos altamente complexos,

dispendiosos e difíceis de operar para quantificar o uso da água pelas plantas e consequentemente gerir os eventos de rega necessários a uma boa produção.

### 2.2.1 – A evapotranspiração e a equação de Penman-Monteith

As necessidades hídricas de uma cultura são definidas como a quantidade de água necessária para suprir a água consumida por uma planta através de evapotranspiração, num meio em que a planta não apresente qualquer restrição ao seu crescimento e sanidade vegetal (Pereira *et al.*, 2013).

A evapotranspiração é o fluxo de água para a atmosfera a partir da superfície do solo (evaporação) e dos estomas presentes nas folhas das plantas (transpiração) bem como da evaporação directa da vegetação quando molhada. Não é fácil distinguir estas componentes uma vez que ocorrem em simultâneo.

Para além da água presente à superfície do solo, o factor que mais contribui para a evaporação é a energia disponível, relacionada com a radiação solar que atinge as primeiras camadas superficiais do solo. À medida que a cultura se desenvolve, o peso da evaporação tende a diminuir.

Existem vários factores que influenciam a ET e que podem ser divididos em factores meteorológicos (radiação, temperatura, humidade e velocidade do vento) e culturais (tipo de cultura, sistema de condução, desenvolvimento cultural, IAF e altura da cultura).

Em agricultura, obter dados fidedignos de ET é o elemento central de qualquer sistema para estimativa de necessidades das culturas, quer na fase de planeamento quer na gestão das explorações.

Existem uma grande variedade de métodos para determinar ET, sendo que a sua utilidade depende dos objectivos dos respectivos estudos. Por exemplo em aplicações em que é necessário prever os valores de ET a sua obtenção será sempre por estimativa a partir de um modelo, sendo por isso conveniente separar os métodos de medição dos de modelação.

Os métodos de medição (directa ou indirecta) da ET podem dividir-se em várias categorias (métodos hidrológicos, micrometeorológicos e métodos baseados na fisiologia das plantas). Qualquer um destes métodos de medição da ET envolvem custos e instalações complexas não sendo viáveis para um uso diário nas operações de condução da rega. São usados maioritariamente em investigação (Pauwels *et al.*, 2006; Conceição *et al.*, 2009; Conceição, *et al.*, 2011, 2017).

No âmbito deste estudo são utilizados dados de ET obtidos a partir da evapotranspiração de referência ( $ET_0$ ) estimada pela equação de Penman-Monteith com parâmetros da relva, sendo este o método recomendado pela FAO, a que se aplicam coeficientes para obter o valor real da ET ( $ET_a$ ).

Criada em 1948, a equação de Penman combina os métodos do balanço energético e da transferência de massa, derivando uma equação que permite calcular a evaporação de uma superfície de água exposta, a partir de dados meteorológicos de registo habitual. A partir deste

método base, diferentes investigadores, como Monteith (1965), foram estendendo a sua utilização a superfícies de folhas de diferentes plantas através de factores de resistência da superfície, dando origem a uma equação geralmente conhecida como de Penman-Monteith.

Para a utilização desta equação é necessário conhecer a resistência aerodinâmica ( $r_a$ ) e a resistência de superfície ( $r_s$ ).

Ambas estas resistências podem ser estimadas considerando as relações entre a copa e as condições meteorológicas na sua envolvência, sendo que para tal é necessário assumir pressupostos na sua interação e interdependência, o que torna a sua utilização extramente difícil para estimar ET directamente (Pereira *et al.*, 2013).

A utilização da forma geral desta equação (Equação 1) requer o conhecimento das resistências, sendo da forma:

$$\lambda ET = \frac{\Delta(R_n - G) + \rho_a C_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma(1 + \frac{r_s}{r_a})} \quad (1)$$

em que  $R_n$  é o balanço das trocas radiactivas, ou radiação líquida,  $G$  é fluxo de calor para o solo,  $(e_s - e_a)$  representa o défice da pressão de vapor,  $\rho_a$  é a densidade média do ar a pressão constante,  $C_p$  é a capacidade térmica mássica do ar,  $\Delta$  o declive da curva de pressão de vapor saturante em função da temperatura,  $\gamma$  a “constante” psicométrica e  $r_s, r_a$  as resistências em fase vapor (Monteith, 1965)

Na verdade, os processos de trocas num coberto vegetal são demasiado complexos para serem apenas representados por estas resistências, como descrito sobre os pressupostos necessários à obtenção de estimativas, mas a sua utilização permite obter boas correlações com as medições, se considerarmos uma superfície uniforme, em especial uma relva.

Ora, por várias razões a  $ET_0$  tinha sido definida décadas antes para uma superfície de referência, um coberto continuo de relva uniforme. Assim, Allen et al. (1998) propuseram o uso da equação de Penman-Monteith com parâmetros da relva, isto é, valores fixos para as resistências em fase vapor, de uma relva em condições de referência, o que representa um caso particular da equação geral proposta em 1965.

Tendo em conta estes parâmetros a equação de Penman-Monteith é dada pela seguinte equação (Equação 2) (Allen et al., 1998):

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (2)$$

em que  $u_2$  é a velocidade do vento em m/s medida a 2 metros de altura e as outras grandezas têm o significado dado na equação 1.

## 2.2.2 – Coeficiente cultural e coeficiente de stress hídrico

Após obter o valor de  $ET_0$ , há que chegar aos valores da evapotranspiração cultural ( $ET_c$ ) e da evapotranspiração real ( $ET_a$ ). Em primeiro lugar introduz-se um coeficiente que traduz a influência que cada tipo de cultura exerce no valor da ET, em relação a  $ET_0$  e permita assim a estimativa de  $ET_c$  (Allen *et al.*, 1998). Para tal esta influência é expressa através de um coeficiente denominado coeficiente cultural ( $K_c$ ) que tem sido classicamente usado em engenharia da rega, obtendo-se  $ET_c$  pelo produto de  $ET_0$  por  $K_c$ . Note-se que o valor de  $ET_c$ , segundo normas internacionais e a praxis na Europa e na maior parte do mundo, se reporta a condições de conforto hídrico (Perrier, 1985).

Os métodos de medição hidrológicos podem ser utilizados para medir directamente  $ET_c$ . No entanto, como já referido acima, para finalidades de aplicação e projecto, estima-se  $ET_c$  a partir de dados meteorológicos e da cultura com a equação de Penman-Monteith e os coeficientes culturais. Em suma, a influência das variáveis meteorológicas é traduzida na estimativa de  $ET_0$  ao passo que o valor de  $K_c$  é responsável por representar as diferenças entre a cultura de referência (relva) e a cultura em estudo, diferenças essas relacionadas com o seu desenvolvimento, características morfológicas e fisiológicas.

Existem duas formas de apresentar o coeficiente cultural, como uma só variável, ou como a soma de dois componentes,  $K_c = K_{cb} + K_e$  em que  $K_{cb}$  representa as características da planta e o seu efeito na transpiração da mesma, e  $K_e$  a componente evaporativa do solo (Allen *et al.*, 2005). Esta segunda forma é de especial relevância para culturas perenes em que existe uma grande parte do solo exposto à radiação solar e que contribui bastante para a evaporação total do coberto vegetal.

Como é de prever, a espécie, seja uma herbácea ou lenhosa, influencia bastante os valores de  $K_c$ , bem como as constantes mudanças na planta durante o seu ciclo de crescimento (Allen *et al.*, 2009).

Para além disso, e apesar de já ter sido referido que  $ET_0$  traduz a influência das condições meteorológicas no cálculo de  $ET_c$  é importante referir que o clima pode influenciar os valores de  $K_c$  em casos em que a resistência aerodinâmica da planta em estudo difere substancialmente da cultura de referência. O vento é um dos factores que mais pode inflacionar o valor de  $K_c$  sendo a sua influência especialmente acentuada em climas áridos (Doorenbos, J. and Pruitt, 1977).

Os valores de  $K_c$  normalmente utilizados na gestão de rega são obtidos a partir de valores tabelados, por exemplo, no manual conhecido como FAO 56 (Allen *et al.*, 1998) ou obtidos experimentalmente para determinadas zonas específicas e constantes da literatura. Não existem valores de  $K_c$  universal que se adapte a todas as situações climáticas e de desenvolvimento cultural, fazendo com que a utilização deste coeficiente na estimativa de  $ET_c$  esteja sempre

rodeado de uma incerteza significativa, especialmente em espécies lenhosas. Na figura 3 apresenta-se a evolução do valor de  $K_c$  ao longo do ciclo cultural.

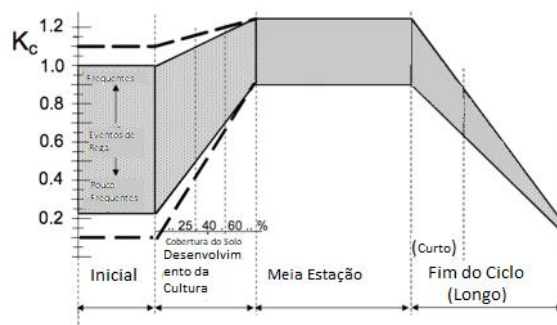


Figura 3 - Evolução de  $K_c$  ao longo do ciclo cultural (Allen et al., 1998).

Esta figura é um exemplo de típico em que se divide o período de crescimento em quatro fases distintas, tentando assim englobar a dinâmica do crescimento da planta ao longo de um ciclo cultural e onde se apresentam os principais factores que influenciam o valor de  $K_c$  ao longo do ciclo, sendo numa fase inicial influenciado pela importância da componente evaporação do solo, depois pela cobertura do solo decrescente com o desenvolvimento rápido da planta e posteriormente pela arquitectura e outros aspectos da cultura (altura final, largura da copa, etc), como descrito por exemplo por Jensen *et al.*, 2016.

Existem vários estudos que permitem correlacionar os valores de  $K_c$  com a medição de outras variáveis como por exemplo o índice de vegetação por diferença normalizada (IVDN) (Campos *et al.*, 2010; Ferreira *et al.*, 2011), ou a acumulação de GD's (Kullberg *et al.*, 2017), apesar de serem baseados em relações empíricas estudadas para culturas específicas, permitem definir com maior certeza a evolução dos valores ao longo do ciclo cultural nos raros casos estudados sendo então uma clara evolução face à consulta de valores tabelados (note-se porém que entre a consulta de tabelas e a modelação de  $K_c$  há uma vasta gama de possibilidades).

No clima mediterrânico é muito frequente encontrarmos zonas em que a escassez de água estival faz com que haja uma limitada disponibilidade de água no solo para absorção pela raiz das plantas, em condições de sequeiro, sendo esta um factor limitante no seu crescimento.

Como tal, nessas condições de sequeiro ou em condições de rega deficitária, é necessário proceder ao ajuste de  $ET_c$  por um factor adicional, o coeficiente de stress,  $K_s$ .

Este coeficiente tem tendência a ser usado com cada vez mais frequência devido às previsões sobre a disponibilidade da água já discutidas no capítulo 2.1. e à consequente importância crescente da rega deficitária, em culturas tradicionalmente de sequeiro (olival, vinha, pomares tradicionais de figueiras, alfarrobeiras, amendoeiras, pereiras, etc) mas também em outros pomares habitualmente regados.

Quando o potencial energético da água presente no solo ou outro indicador de stress hídrico usado, baixa para lá de um nível crítico, podemos considerar a cultura como estando sob stress hídrico. A equação para  $ET_a$  é descrita por (Equação 3):

$$ET_a = K_s K_c ET_0 \quad (3)$$

Allen et al., (1998) consideram que  $K_s$  afecta apenas a transpiração das plantas, e assim apresentam a equação na forma seguinte (Equação 4) que, por não ser consensual em termos de notação, não usaremos:

$$ET_c = (K_s K_{cb} + K_e) ET_0 \quad (4)$$

É, com efeito, cada vez mais comum a gestão de rega considerar períodos em que a cultura passa por um período de stress hídrico (rega deficitária). Este tipo de gestão tem demonstrado não só reduzir os consumos de rega como melhorar factores de qualidade, como o grau brix e outros, o que também pode contribuir para a melhoria do rendimento do agricultor (Lorite et al., 2007).

Existem várias técnicas para a rega deficitária como a rega deficitária sustentada, rega deficitária regulada e a rega alternada das raízes (regando apenas um lado da linha e alternando entre os lados (Geerts et al., 2009).

### 2.2.3 – Balanço hídrico

Para uma boa gestão de rega é necessário relacionar as necessidades hídricas da cultura com a quantidade de água disponível no solo para a cultura diariamente.

A estimativa correcta destas duas variáveis permite fazer a verificação do balanço hídrico do solo, onde consideramos que o solo é um volume de controlo que recebe e perde água pelos processos exemplificados na figura 4 com setas que representam entradas e saídas.

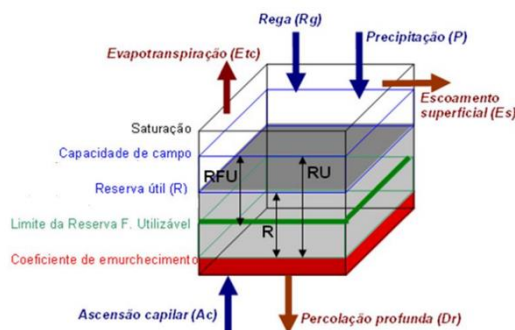


Figura 4 – Processos envolvidos no balanço hídrico do solo. Adaptado de (Allen et al., 1998).

Para a condução da rega, independentemente da frequência (diária, semanal ou planeamento) é necessária a compreensão do significado do teor em água no solo e, assim,

começa-se por obter informação acerca do teor de água à capacidade de campo (CC) e ao coeficiente de emurchecimento (CE).

O teor de água à capacidade de campo ( $\theta_{cc}$ ,  $m_3/m_3$ ) é o teor de água que permanece num determinado solo após a drenagem e escoamento superficial da água em excesso e a partir do momento em que o movimento descendente da água na unidade de solo considerada cessa praticamente, o que pode levar entre um a três dias após um evento de rega ou precipitação (Veihmeyer *et al.*, 1949). O solo continua a drenar a água existente no perfil, mas o seu movimento é tão lento, quando comparado com a taxa de extração de água pelas plantas, que pode ser ignorado. É medido na grande maioria dos casos como sendo a massa de água por unidade de massa do solo seco. O seu valor varia de solo para solo, podendo ser determinado experimentalmente *in situ*, por determinações laboratoriais para determinada pressão negativa (menos preciso) ou ainda podendo ser estimado muito grosseiramente a partir de uma análise granulométrica do solo.

O teor de água no coeficiente de emurchecimento ( $\theta_{ce}$ ,  $m_3/m_3$ ) é o valor que representa o nível de água presente no solo a partir do qual a planta sofre danos irreversíveis (emurchecimento permanente)

Considera –se que a reserva útil de água no solo (RU, mm ou m, unidade de z) é dada pela expressão (Equação 5):

$$RU = (\theta_{cc} - \theta_{ce}) * z \quad (5)$$

sendo que z representa a profundidade do sistema radicular, no seu limite inferior, isto é o que efectivamente delimita a zona do solo em que as raízes da planta se encontram e absorvem água.

Para obter realmente a quantidade de água facilmente disponível para a planta é necessário ter em consideração um índice, que de aqui em diante designaremos por défice permissível (p), e que representa a fração de água facilmente utilizável em relação à reserva utilizável ou útil. Este factor é necessário pois antes de atingir o coeficiente de emurchecimento, a capacidade da planta de obter água do solo vai-se reduzindo significativamente, a partir deste nível crítico. Através deste factor é então possível obter a reserva de água facilmente utilizável pelas plantas (RFU) (Equação 6):

$$RFU = RU * p \quad (6)$$

é importante realçar que p não deve ser considerado como um coeficiente escolhido por razões de economia no uso da água disponível, mas sim como uma variável estritamente relacionada com o sistema eco fisiológico formado entre a planta e o solo, representando o limiar a partir do qual a planta deixaria de estar em conforto hídrico, havendo redução de ET (Ferreira, 2017). A estimativa da água facilmente disponível a partir deste método empírico tem vindo a ganhar



apoiantes na comunidade científica, que consideram este método fácil e com estimativas aceitáveis para o seu uso como indicador (Turner, 1986; Passioura, 1988).

O valor de  $p$  não é estático, dependendo entre vários outros factores do poder evaporativo da atmosfera e em certa medida, da textura do solo e distribuição radicular (Sadras *et al.*, 1996). Allen *et al.* (1998) apresentam valores tabelados para condições edafo-climáticas consideradas normais e uma fórmula para estimar este valor em condições de  $ET_c$  díspares.

O balanço hídrico pode ser feito para obter como incógnita a variação de armazenamento  $\Delta S$  igual à variação entre o armazenamento de água no solo, por exemplo no próprio dia ( $S_i$ ) e no dia anterior ( $S_{i-1}$ ), variação essa estimada a partir das entradas e saídas de água no volume de solo considerado, para um período de tempo determinado (geralmente um dia), podendo ser traduzida pela seguinte expressão (Equação 7):

$$\Delta S = P_e - ET_c + R - ES + A_c - D_r \quad (7)$$

em que  $P_e$  é a precipitação efectiva,  $R$  a dotação total de regas efectuadas,  $ES$  o escoamento superficial,  $A_c$  a ascensão capilar e  $D_r$  a percolação profunda ou drenagem.

Tipicamente e para efeitos de condução da rega ignoram-se  $ES$ ,  $A_c$  e  $D_r$ , tem-se  $P_e$ ,  $R$  e  $ET_c$  respectivamente medidos e estimado como acima descrito, obtendo-se  $\Delta S$  da equação 7.

Porém, pode verificar-se se esse valor corresponde à realidade por métodos directos de amostragem para obter o valor de  $\theta$ , comparação essa que permite analisar a fiabilidade da estimativa já que  $S$  é igual a  $\theta(m_3/m_3) \times z$ .

De realçar também que o valor de  $K_s$  diário em situações de stress hídrico, ou seja, situações em que o valor da reserva do solo está abaixo do limite inferior da RFU, pode ser obtido a partir da relação entre esse valor e o limite da RFU, como baixo indicado (Equação 8):

$$K_s = \begin{cases} 1, & \text{se } R_d < RFU \\ \frac{RU - R_d}{RU - RFU}, & \text{se } R_d > RFU \end{cases} \quad (8)$$

em que  $R_d$  é o somatório de  $ET_a$  desde o último evento de rega.

Esta é a forma mais fácil e expedita de calcular o valor de  $K_s$  nestas condições. No entanto é de realçar que factores e características de cada solo podem influenciar este valor como os referidos acima e discutidos em Ferreira (2017), onde se referem as limitações desta abordagem em olival.

A partir dos valores de RFU e  $ET_c$  é possível determinar o intervalo máximo entre regas e a dotação de rega máxima, para cada situação, valores muito usados quer na gestão da rega, quer em projecto de rega. Quando a rega repõe o solo à CC, a dotação de rega é igual ou inferior ao seu valor máximo, sendo a dotação sempre igual à diferença entre o estado da reserva antes da

aplicação da rega e o valor correspondente à CC, ambos expressos em volume por unidade de área (m ou mm).

## **2.3 – Indicadores do estado hídrico**

Face às metodologias apresentadas ao longo do subcapítulo 2.2 é possível ao agricultor com o auxílio de dados simples e a determinação das características do solo e da cultura obter todos os elementos necessários a uma gestão de rega minimamente fundamentada. No entanto e face à crescente competitividade do sector agrícola, a escassez de recursos e as restrições e normas a cumprir, a condução da rega tem de ser cada vez mais precisa e aproximações grosseiras deixam de ser suficientes para uma gestão eficiente. Como tal é cada vez mais relevante avaliar em tempo real a dinâmica da água no solo e as mudanças morfológicas e fisiológicas da planta. Com este objectivo em mente, existem no panorama actual várias tecnologias que são já auxiliares importantes à gestão da rega e que cada vez mais fazem parte do leque de ferramentas do dia-a-dia para a condução de rega.

### **2.3.1 – Indicadores do estado hídrico do solo – Sondas**

Quantificar precisamente a quantidade de água retida numa determinada unidade de volume de solo é uma tarefa praticamente impossível a não ser que controlemos completamente o ambiente como no caso de um lisímetro em ambiente experimental. Na realidade obtêm-se estimativas baseadas em amostras recolhidas ou medições pontuais ao longo do tempo.

Conforme abordado no início deste subcapítulo, esta informação sobre o solo não é apenas relevante para a condução da rega, mas afecta processos como a germinação de sementes, o seu crescimento, a sua nutrição e até a degradação microbiológica da matéria orgânica.

Por isso mesmo, ao longo dos anos têm sido desenvolvidos métodos e sensores para medir ou estimar o conteúdo de água no solo a diferentes escalas temporais e espaciais. Como existem uma grande quantidade de metodologias, neste trabalho irão apenas ser expostas as principais e detalhadas as mais relevantes nos dias que correm (Bittelli, 2011).

É importante especificar alguns dos princípios físicos em que baseiam e que permitem obter medições directas ou indirectas. Nas medições directas a quantidade de água é medida, por exemplo, medindo o seu peso como uma fracção do peso total da amostra, denominando-se este método por método gravimétrico. É o método mais tradicional e utilizado em ambiente científico permitindo as medições mais fidedignas. No entanto é um método invasivo que destrói a estrutura do solo em análise, é muito trabalhoso e pouco prático (Shukla *et al.*, 2014) sendo usado para calibração de sensores por outras abordagens.

Foram aparecendo alternativas de medição indirecta que permitem medir uma outra variável que seja afectada pelo conteúdo de água no solo, sendo posteriormente associada por

relações empíricas ou físicas denominadas curvas de calibração. O quadro 2 resume estes métodos e a suas principais características.

Quadro 2 – Métodos de determinação da humidade do solo (Vienken et al., 2013).

<b>Tipo</b>	<b>Nome</b>	<b>Detalhes</b>
<b>Medição Directa</b>	Gravimetria	<ul style="list-style-type: none"> <li>- Método de referência</li> <li>- Determinação do peso de água em amostras antes e depois de um processo de secagem em estufa.</li> <li>- Considera-se a diferença entre os dois valores, em relação ao peso seco.</li> <li>- Utilizado muitas vezes para calibrar sensores.</li> </ul>
<b>Medição Indirecta</b>	Sonda de Neutrões	<ul style="list-style-type: none"> <li>- São emitidos neutrões a partir de uma fonte de radiação.</li> <li>- Perdem energia ao colidir com outros átomos, especialmente os de hidrogénio.</li> <li>- Após esta perda energética, alguns sofrem alterações na trajectoria e são reflectidos na direcção da fonte, onde são contados pelo detector.</li> <li>- As contagens são transmitidas e processadas num microprocessador por uma equação de calibração.</li> <li>- O equipamento é constituído por um emissor e um detector</li> <li>-Método indirecto mais preciso</li> </ul>
	Resistividade	<ul style="list-style-type: none"> <li>- Corrente transferida para o solo por eléctrodos</li> <li>- A resistividade é medida pelas alterações na voltagem</li> </ul>

		- Usado em mapas de condutividade eléctrica
	Dielétricos	<ul style="list-style-type: none"> <li>- Avaliam as diferenças nas constantes dieléctricas das diferentes componentes do solo (sólidos, líquidos e gás).</li> <li>- É muito sensível a variações do conteúdo de água</li> <li>- Obtém-se o volume através de curvas de calibração</li> </ul>

Dos métodos apresentados, os mais utilizados actualmente em ambiente comercial são os indirectos, com especial ênfase nos que avaliam a constante dieléctrica do solo, tendo em consideração que a sonda de neutrões envolve perigos radiológicos e tem custos elevados.

Existem sensores deste tipo com dois tipos de princípio de detecção. Os TDR (*Time-Domain- Reflectometer*) cujo princípio de funcionamento inclui uma ponta de prova que mede a velocidade de ondas eletromagnéticas que viajam entre dois pontos no solo, sendo possível quantificar o tempo que a onda leva a percorrer este percurso e inferir o conteúdo de água naquele espaço por relações empíricas (Topp *et al.*, 1985; Robinson *et al.*, 1999).

Os FD (*Frequency Domain*) ou sensores de capacitância medem directamente a constante dieléctrica, detectando alterações na capacitância do solo com alterações de conteúdo em água; são mais fáceis de construir que os TDR e apresentam tempos de resposta mais rápidos. Necessitam de ser calibrados com precisão devido à complexidade dos campos eléctricos que geram (Miller *et al.*, 2014; Shukla *et al.*, 2014).

A utilização destes sensores é hoje em dia uma realidade na maior parte das explorações comerciais. O seu uso abrangente leva a que cada vez mais as soluções apresentadas no mercado sejam cada vez mais precisas e que já permitam medidas aproximadamente correctas mediante uma boa calibração. Por outro lado, a sua sensibilidade é por vezes suficiente para tirar ilações acerca da dinâmica do solo e da sua resposta às entradas e saídas de água, o que já de si é uma grande mais valia (Vera *et al.*, 2009).

### 2.3.2 – Indicadores do estado hídrico da planta – Dendrómetros

O sistema solo-planta não pode ser encarado como composto por partes independentes, uma vez que todas as alterações sofridas no solo se reflectem na planta. Este tipo de relação implica que seja necessário não só conhecer o conteúdo de água no solo, mas também obter informação sobre o estado hídrico da planta. A calendarização da rega baseada em medições recolhidas da planta tira partido desta relação, funcionando a planta como um biossensor, refletindo até certa medida o estado hídrico da mesma (Goldhamer *et al.*, 2001).

Para obter estes indicadores da planta é necessário considerar elementos na morfologia da planta que possam reflectir o seu estado hídrico, seja por medição directa ou por relações já estudadas, tendo sempre em consideração que a adequabilidade de cada indicador varia consoante o tipo de cultura e as suas características.

Existem inúmeros sensores cuja principal aplicação é precisamente inferir sobre os movimentos e fluxos de água na planta, como a câmara de pressão (potencial foliar ou do ramo) ou variações da temperatura da copa. Estes métodos apresentam validade experimental e a sua utilização conduz a bons resultados práticos, dependendo dos casos. No entanto não são automatizáveis, ou a sua automatização apresenta custos elevados inoportáveis para o produtor. Daí que métodos como a medição da variação do diâmetro do tronco (dendrometria) a intervalos regulares seja uma alternativa relativamente barata, por os sensores serem fáceis de monitorizar e automatizar.

Neste trabalho abordamos apenas a dendrometria devido à sua simplicidade comparada com a complexidade e impossibilidade de automação das câmaras de pressão.

Para perceber como a flutuação do diâmetro da planta está associada ao seu estado hídrico consideremos o fluxo de seiva. A abertura estomática das folhas conduz à criação de uma tensão entre a superfície evaporativa das folhas e os órgãos acumuladores de água na planta. O diâmetro do tronco decresce como consequência, devido ao movimento ascendente de água em direcção às folhas, provocando variações mais ou menos acentuadas, dependendo da elasticidade dos tecidos e do tipo de planta. Esta perda de água é facilmente revertível, ou mesmo evitada se existir um bom teor de água no solo que permita à planta rehidratar-se. A rehidratação contraria os efeitos da perda de água por transpiração e evaporação superficial. As variações do tronco, quando presentes, são medidas normalmente em micrómetros (Ginestar *et al.*, 1998; Zweifel *et al.*, 2001, 2005).

Actualmente os sensores mais utilizados em dendrometria são do tipo LVDT (transformadores diferenciais de variabilidade linear). Estes sensores são robustos e de alta precisão, próximo de  $\pm 1$  micrómetro. No entanto, a precisão real no campo, muitas vezes pode ser da ordem dos  $\pm 10$  micrómetros porque podem existir erros de calibração, ou mudanças bruscas de temperatura. No processo de instalação alguns modelos são pregados ou aparafusados nas árvores, quer seja para segurar o instrumento, quer seja para servir de referência fixa para medição do crescimento e são instalados no lado oposto da trajectória do sol para diminuir efeitos de aquecimento. Silva (2009) apresenta uma discussão sobre os cuidados de instalação.

A interpretação dos diâmetros medidos é um desafio complexo pois, como já foi referido existem plantas cujos ciclos de recarga e perda de água se reflectem com clareza em alterações de diâmetro, enquanto que outras, por terem mecanismos internos de regulação não são tão sensíveis e torna-se difícil inferir qualquer tipo de conclusão sobre o seu estado hídrico. No entanto, se calibrados e com sensibilidade suficiente é possível encontrar padrões e ciclos de crescimento dos diâmetros que estão relacionados com alterações dos ciclos de rega, por exemplo. Alguns estudos indicam que é possível realizar toda a gestão de rega apenas com base

nos dados de dendrometria (Ginestar *et al.*, 1998; Fernández *et al.*, 2010; Corell *et al.*, 2014) enquanto outros descartam essa hipótese (Carr, 2013), mas os resultados concretos dependem da variável derivada que seja utilizada. Pode ser uma fonte de dados interessante da qual podemos retirar vários índices de estudo e avaliação, especialmente em culturas lenhosas.

Existem vários problemas a superar quando se pretende utilizar a dendrometria como indicador do estado hídrico. Muitas vezes as alterações no diâmetro do tronco são influenciadas por outros factores igualmente importantes como a quantidade de frutos na planta, o seu tamanho, idade e ciclos de crescimento intrínsecos despoletados por características varietais e o valor de  $ET_0$ . Existem também outros factores como o tamanho do câmbio ou a densidade de plantação. Como exemplo, num estudo em oliveiras (*Olea europaea*) da variedade “Manzanilla” com cerca de 40 anos concluiu-se que a utilidade de índices derivados de medições do diâmetro do tronco é praticamente nula em anos de safra (Girón *et al.*, 2016). No entanto em casos de stress hídrico significativo estas medições detectaram o stress, podendo ser utilizadas como alertas para situações extremas de stress na planta. Outros estudos sugerem que, em plantas jovens em que o tronco possui um diâmetro mais pequeno a utilização destes índices derivados na condução diária da rega é possível (Goldhamer *et al.*, 2001; Naor *et al.*, 2003; McCarthy *et al.*, 2014).

Em culturas florestais existe a noção de que grande parte do crescimento dos troncos, processo fundamental à valorização da madeira ocorre devido principalmente a eventos de precipitação. Claro que em culturas anuais ou perenes esta avaliação depende dos factores considerados acima, mas este tipo de afirmação contribui para enfatizar a importância destas medições se bem processadas.

No quadro 3 apresentam-se os principais índices derivados das medições de diâmetro do tronco, embora noutros estudos no país se tenham, por necessidade, explorado outras alternativas na falta destas (Ferreira *et al.*, 2012; Silva *et al.*, 2014) :

Quadro 3 – Principais índices derivados da dendrometria (Fernández *et al.*, 2010).

Nome	Expressão de cálculo
<b><i>MDS</i></b>	$MXTD_d - MNTD_d$
<b><i>DR</i></b>	$MXTD_{d+1} - MNTD_d$
<b><i>TGR</i></b>	$MXTD_{d+1} - MXTD_d$
<b><i>Signal Intesity</i></b>	$MDS_{exp} / MDS_{ref}$

A partir destas expressões conseguimos obter valores que podem ou não ter relações com o nível de stress da planta. Os índices apresentados são computados a partir das leituras dos dendrómetros sendo, “Maximum Daily Shrinkage” (*MDS*) obtido pela diferença entre o maior valor diário registado ( $MXTD_d$ ) e o menor ( $MNTD_d$ ), “Daily Recovery” (*DR*) obtido pela diferença entre o valor máximo do dia seguinte ( $MXTD_{d+1}$ ) e o mínimo do dia actual ( $MNTD_d$ ), sendo “Trunk

*Growth Rate*” ( $TGR$ ) obtido pela diferença entre o valor máximo do dia seguinte ( $MXTD_{d+1}$ ) e o máximo do dia actual ( $MXTD_d$ ).

O valor mais representativo é obtido pelo índice designado por alguns autores de “*Signal Intensity*” ( $SI$ ), usado em Ferreira *et al.*, 1996, com a designação de amplitude diária relativa (RDTS, *Relative daily trunk shrinkage*), sendo que para a sua obtenção é sempre necessário ter em consideração um valor de referência para um mesmo dia a partir da medição do tronco de uma planta em conforto hídrico ( $MDS_{ref}$ ). Este detalhe impede que este índice de maior sensibilidade possa ter valor num ambiente comercial, pois é impossível ter certezas sobre o estado hídrico dessa referência e muitas vezes é incomportável ter dois pontos de medição. Por esse motivo é necessário desenvolver novas metodologias que permitam ser tão sensíveis como o IS mas que possam ser extrapoladas ou inferidas a partir de outro conjunto de variáveis auxiliares (Fernández *et al.*, 2010) o que se tem relevado uma tarefa ainda não alcançada na generalidade das situações (Carr, 2013).

## **2.4 – IoT, Big Data e agricultura de precisão**

### **2.4.1 – Um mundo de ligações**

O avanço tecnológico e comunicacional dos últimos anos impulsionou o desenvolvimento de novas ferramentas que são hoje em dia utilizadas um pouco por todos os ramos da sociedade e nas indústrias.

O conceito de IoT nasce devido ao aumento da cobertura e largura de banda disponível para ligações móveis e dos progressos nas tecnologias de computação, com chips cada vez mais pequenos e potentes. A combinação destes dois factores levou a que fosse possível implementar redes sensoriais em qualquer ambiente desde as casas inteligentes até à monitorização em tempo-real dos sistemas de rega.

A proliferação de sensores compatíveis com um grande conjunto de microchips e a expansão de projectos de baixo custo como o *arduino* ou o *raspberry pi* permitem a qualquer pessoa criar os seus protótipos e construir redes de monitorização para as mais variadas finalidades. No panorama actual estima-se que existam 23,14 milhões de dispositivos ligados em rede, com a tendência de um elevado crescimento na próxima década. Na figura 5 podemos observar a tendência de crescimento no número de dispositivos conectados.

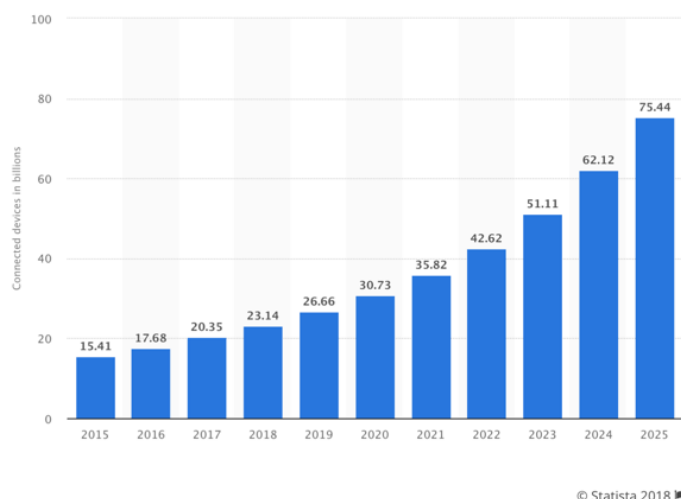


Figura 5 – Número de dispositivos conectados em rede e projeção até 2025.

Como foi abordado no capítulo anterior, os sensores para aplicações na agricultura são uma realidade incontornável e uma ferramenta de apoio à decisão com ampla utilização, vindo a sua implementação a acompanhar o crescimento da IoT. É óbvio que este conceito de equipamentos ligados em rede, onde é possível ver num smartphone, tablet ou PC, dados de plantas, solo, meteorologia, presença de patogénicos ou pragas, em tempo real e em qualquer lugar do planeta, dá-nos a sensação de transportar a agricultura para o século XXI, entrando num mundo mais tecnológico, embora na raiz não possam deixar de estar os conhecimentos relevantes sobre os processos em curso. Existem já muitas aplicações e empresas a explorar estes conceitos e a desenvolver produtos que apresentam resultados concretos em termos de produção, uso eficiente dos recursos e prevenção de pragas e doenças (Sonka, 2014).

#### 2.4.2 – Dados e o seu potencial para criar valor

Os dispositivos de IoT referenciados no subcapítulo anterior geram um crescente volume de dados. Todos estes dados acumulados são referenciados pelo termo “*Big Data*” devido à sua complexidade de processamento pelas ferramentas tradicionais. Os dados considerados “*Big Data*” podem ser caracterizados por 3 V’s: Volume, Velocidade e Variedade. Segundo os dados estatísticos mais recentes é expectável que em 2020 sejam gerados 1.7 MB/s por cada ser humano. O ponto mais interessante destas características é a variedade, pois hoje conseguimos obter estes dados de inúmeras fontes e de diversos tipos, dados de GPS, fotos, imagens aéreas de drone ou de satélite, etc. Este crescente número de dados diferentes apresenta potencialidades únicas para análise, permitindo descobrir relações e valor acrescentado onde anteriormente apenas conseguíamos obter ruído (Sonka, 2014; Kamilaris *et al.*, 2017).

Para explorar todo o potencial do “*Big Data*” na agricultura é necessário fazer um pré-processamento cuidadoso, exigir bastante critério na aquisição de dados e adaptar unidades de forma a tornar a análise um processo válido do ponto de vista agronómico, uma vez que as relações estabelecidas têm de ter significado à luz dos processos físico-químicos e biológicos.



Cada vez mais existem estudos relacionados com a exploração de conjuntos de dados usualmente designados de “*Big Data*”. O mundo agrícola já entende o valor destes dados, existindo plataformas onde é possível comprar e vender dados directamente dos produtores. Um ponto chave nesta partilha de dados é o aparecimento de plataformas como o *Google Cloud* ou a *Amazon Web Services* que permitem armazenar e aceder aos dados em qualquer ponto do mundo, a partir dos seus serviços na *cloud*.

### **2.4.3 – A agricultura de precisão**

A agricultura de precisão envolve a integração de novas tecnologias como a detecção remota, a recolha de dados de colheita e mais recentemente o uso de drones, entre outras para permitir ao agricultor fazer uma gestão das culturas em que é possível avaliar e qualificar diferentes características da cultura e do solo dentro das parcelas, introduzindo a possibilidade de criar zonas de gestão com variabilidade espacial muito maior, o que conduz a um tratamento e condução da cultura muito mais personalizado às necessidades de cada ponto.

A ascensão da agricultura de precisão começou com o aparecimento dos primeiros sistemas de GPS que foram a oportunidade de cartografar e georreferenciar pontos de interesse ou áreas com diferente vigor vegetativo, introduzindo o conceito de variabilidade espacial dentro das parcelas. Este conhecimento foi sendo melhorado e aprofundado com o uso de imagens multiespectrais, a colocação de sensores nas máquinas agrícolas e mais recentemente a avaliação do teor de humidade do solo e indicadores hídricos das plantas. A ampla adopção destas tecnologias suportou a introdução de novas técnicas na gestão das culturas, sendo o exemplo mais comum as técnicas de VTA (Variação da taxa de aplicação) em semeadores e distribuidores de adubo.

Este trabalho foca-se na rega, que por sua vez, também nalguns casos pontuais usufruiu de optimizações com a maior resolução espacial disponível. O objectivo da rega de precisão é regular os débitos de água permitindo adequar os mesmos às necessidades de cada zona homogénea definida. Este aumento na resolução espacial da informação permite à rega de precisão maximizar a relação benefício/custo das operações de rega e os seus mecanismos, que por vez representam grande parte da despesa total do agricultor. Permite ainda adaptar a rega à variabilidade dos solos e do desenvolvimento da cultura (Capraro *et al.*, 2008; Veraa *et al.*, 2017).

## **2.5 – Machine Learning, conceitos e principais algoritmos**

A ampla implementação de redes de sensores IoT e recolha de dados “*Big Data*” no meio agrícola estão a proporcionar uma oportunidade para que tecnologias disruptoras, e que à primeira vista não teriam grande relevância ou possibilidade de implementação na agricultura comum, comecem a ser relevantes na investigação e utilização quotidiana. Com o número de dados recolhidos a aumentar, sobre vários órgãos e mecanismos das plantas, do solo e de todo

o ecossistema em seu redor, estão reunidas as condições para a aplicação de técnicas de *Machine Learning* (ML) na agricultura.

Este é um campo bastante amplo, com algoritmos para a resolução de uma grande variedade de problemas. Neste momento existem já várias aplicações comerciais, sendo um bom exemplo, o *lettuce bot*, que consegue identificar as infestantes e proceder à aplicação de produtos fitofarmacêuticos apenas nessas mesmas infestantes, reduzindo o risco para a saúde do consumidor e poupando recursos e o ambiente (Liu *et al.*, 2010; Gonzalez-Sanchez, 2014; Karandish *et al.*, 2016; Adnan *et al.*, 2017). São tecnologias deste género que só emergem agora como viáveis, e que podem revolucionar o sector.

Um dos grandes desafios é encontrar, ou desenhar o algoritmo adequado a cada objectivo a partir das variáveis disponíveis e garantir a integridade dos dados em ambientes não experimentais.

A possibilidade de encontrar um, ou vários algoritmos que consigam modelar e prever o comportamento das plantas nas mais variadas condições edafoclimáticas irá fornecer ao agricultor uma plataforma onde será possível testar cenários e prever consequências antes da tomada de decisão. Este salto seria um avanço tremendo nas práticas agrícolas actuais e na procura da sustentabilidade.

Existem várias definições do que deve ser considerado como ML. A mais aceite e abrangente é que ML é a ciência responsável por criar algoritmos que permitam a um computador tomar decisões sem serem directamente programadas para aquela tarefa e a melhorar a sua aprendizagem autonomamente, consumindo, para tal, mais dados na forma de observações e interações do mundo real. Isto implica um processo demorado, onde é criado um modelo que se ajusta de forma mais ou menos precisa aquela combinação de dados fornecidos como input (Mitchell, 1997).

O número de algoritmos de ML cresce todos os dias com o número de aplicações dos mesmos, sendo normalmente agrupados por tipo de aprendizagem (Aprendizagem supervisionada, aprendizagem sem supervisão e aprendizagem reforçada) ou por tipo de formulação (classificação, regressão, árvores de decisão, redes neuronais, etc).

Independentemente do tipo, todos os tipos de algoritmos de ML consistem em uma representação (conjunto de classificadores), uma avaliação (o objectivo a alcançar) e uma optimização (método de busca ou classificação) (Shalev-Shwartz *et al.*, 2014).

A obtenção de um algoritmo de ML pode dividir-se nos seguintes passos:

- Pré-processamento, onde são tratados e normalizados todos os dados recolhidos.
- Transformações dos dados e classificação de variáveis, se necessário.
- Divisão dos dados em dados de treino e dados de teste.
- Treino do algoritmo com os dados de treino.
- Validação com os dados de teste (*Benchmarking*)
- Repetir os passos anteriores até obter um algoritmo fiável com diversos conjuntos de dados.

O processo de pré-processamento envolve várias etapas que permitem ajustar o conjunto de dados ao correcto funcionamento dos algoritmos escolhidos. Este processo pode ser dividido nas seguintes etapas:

- Limpeza: corrigir ou remover valores anormais para cada variável, preencher lacunas nos dados (a partir da média/mediana dos dados mais próximos temporalmente)
- Formatação: cada algoritmo precisa que os inputs sejam fornecidos num determinado formato de ficheiro dependendo da biblioteca de programação utilizada (Lista, .csv, .txt)
- Amostragem: nem todos os dados recolhidos são necessários ou acrescentam algo ao funcionamento do algoritmo
- Dimensionamento: as variáveis são obtidas em diferentes unidades e escalas sendo necessário que estas sejam normalizadas para uma escala fixa, garantindo que nenhuma se sobrepõe a outra devido a valores absolutos muito elevados.

A selecção dos algoritmos a testar depende muito do tipo de problema abordado. Para uma tarefa de categorização de variáveis discretas como identificar emails (spam/não spam), estamos perante um algoritmo de classificação. Se, pelo contrário, o objectivo é prever o valor de uma variável contínua utilizamos um algoritmo de regressão. Existem ainda os algoritmos de agrupamento para conjuntos de dados, onde a relação entre as variáveis não é conhecida ou inferida directamente. Este tipo de algoritmos serve principalmente para obter algum tipo de relação entre as variáveis que não é visível à partida, ou para proceder a uma redução da dimensionalidade, eliminando variáveis cuja correlação seja muito elevada (Mitchell, 1997). A figura 6 ilustra o fluxo de etapas para um algoritmo de ML.

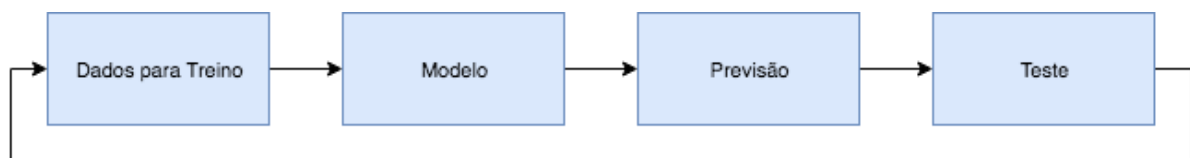


Figura 6 – Fluxo de etapas para desenvolvimento de algoritmos de ML

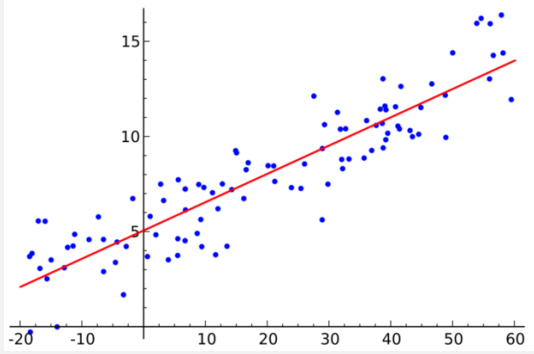
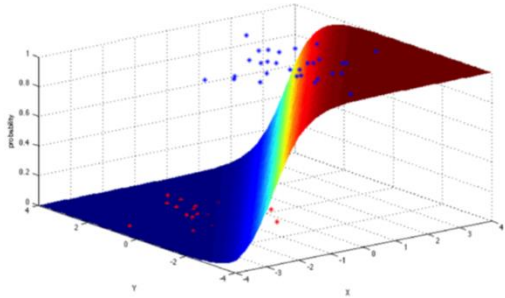
Existem dois problemas fulcrais e recorrentes em ML, o problema da dimensionalidade, onde um algoritmo que considere inúmeras variáveis é mais difícil de interpretar pois apresenta um número maior de dimensões e o problema da sobreestimação, onde um modelo se adapta demasiado às características dos dados de treino, perdendo flexibilidade e capacidade de interpretar correctamente novos dados que sejam muito díspares dos iniciais (Géron, 2017).

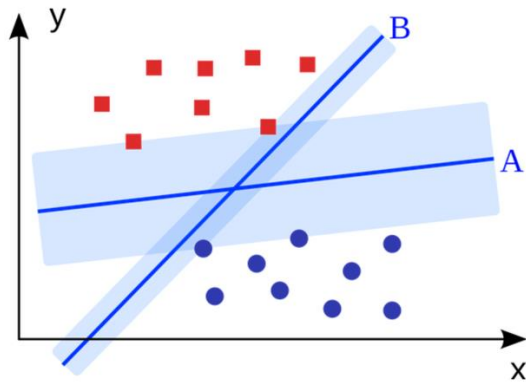
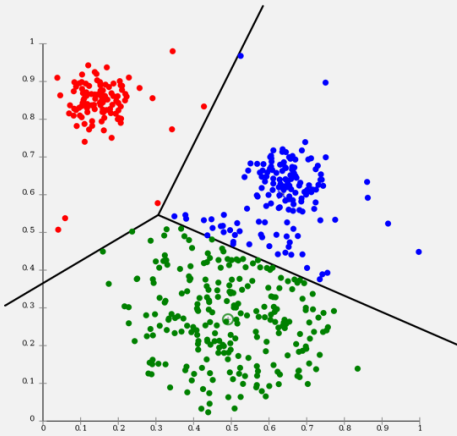
Os elementos que mais contribuem para o sucesso da aplicação de um algoritmo de ML são:

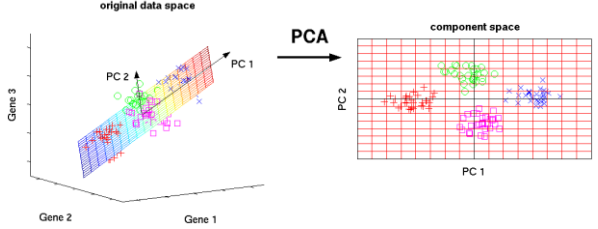
- Escolha de variáveis representativas e independentes;
- Normalização da amplitude e distribuição dos dados;
- Separação dos dados em conjunto de treino e de teste;
- Elevado número de dados disponíveis;

No quadro 4 encontram-se descritos os principais algoritmos tradicionais de ML e uma breve descrição de cada um.

Quadro 4 – Algoritmos comuns de ML (Mitchell, 1997; Shalev-Shwartz et al., 2014; Géron, 2017)

Tipo de aprendizagem	Algoritmo	Descrição
Supervisionada	Regressão Linear 	<ul style="list-style-type: none"> <li>- A análise estatística mais comum.</li> <li>- Utilizada para modelos lineares contínuos</li> </ul> $Y = mx + b$
	Regressão logística 	<ul style="list-style-type: none"> <li>- Aplicação de uma função de transformação, transformando valores contínuos em discretos</li> <li>- Adaptada a classificação binária</li> <li>- Permite obter probabilidades de uma classe padrão</li> </ul> $h(x) = \frac{1}{1 + e^{-x}}$
	Naive Bayes	<ul style="list-style-type: none"> <li>- Calcula a probabilidade de um evento ocorrer, considerando que outro evento já tenha ocorrido.</li> <li>- Calcula-se a probabilidade uma hipótese (h) ser verdadeira, dado o</li> </ul>

Não supervisionada		<p>evento (e) ser também verdadeiro.</p> <ul style="list-style-type: none"> <li>- Assume-se que todas as variáveis são independentes</li> </ul> $P(h/e)$ $= \frac{(P(e/h) * P(h))}{P(e)}$
	<p>SVM</p> 	<ul style="list-style-type: none"> <li>- Classificador binário</li> <li>- Dados dois conjuntos de pontos de dois tipos em N dimensões, SVM gera um hiper-plano de dimensão (N - 1) que separa estes pontos em dois grupos o mais afastados possível um do outro</li> </ul>
	<p>K-means</p> 	<ul style="list-style-type: none"> <li>- Algoritmo iterativo que agrupa dados similares.</li> <li>- Calcula os centróides de cada agrupamento</li> <li>- Atribui um ponto a um agrupamento consoante a sua distância aos centróides (menor distância, ponto pertence ao agrupamento)</li> </ul>
	<p>PCA</p>	<ul style="list-style-type: none"> <li>- Reduz o número de variáveis</li> <li>- Captura a variância máxima nos dados e</li> </ul>

	 <p>The diagram shows the transformation of data from an 'original data space' to a 'component space' using PCA. On the left, the 'original data space' is a 3D plot with axes labeled 'Gene 1', 'Gene 2', and 'Gene 3'. Data points are represented by various colored shapes (circles, squares, triangles). Two principal components, 'PC 1' and 'PC 2', are shown as arrows originating from the origin. An arrow labeled 'PCA' points to the right, where the 'component space' is shown as a 2D plot with axes 'PC 1' and 'PC 2'. The data points are now projected onto this 2D plane, maintaining their relative positions.</p>	<p>constrói um novo sistema de coordenadas</p> <ul style="list-style-type: none"> <li>- Cada componente é uma combinação linear das variáveis originais</li> <li>- São ortogonais em relação um ao outro</li> <li>- A primeira iteração captura a direcção da variável com variância máxima, a segunda a restante variação dos dados</li> <li>- Podem-se encadear análises de PCA, até sobrares apenas duas variáveis (caso extremo)</li> </ul>
--	--	---

Com o uso destes algoritmos é possível obter bons modelos para a maioria das aplicações comuns de ML. No entanto foram deixados de fora desta tabela as redes neuronais e a aprendizagem reforçada, sendo estes dois tipos de algoritmos bastante mais complexos e com outro tipo de considerações. Devido a este factor e à sua importância neste trabalho são apresentados nos próximos dois subcapítulos.

## 2.5.1 – Redes neuronais

### 2.5.1.1 – Introdução aos conceitos e perceptrão

As redes neuronais, como o próprio termo indica, baseiam-se em tentar imitar as funcionalidades dos neurónios do cérebro humano num computador, ou ambiente computacional adequado.

A computação tradicional é perfeita para guardar uma vasta quantidade de informação e reorganizá-la de acordo com instruções precisas criadas especificamente para aquela tarefa e contendo todas as possibilidades e condições da mesma. Por sua vez o cérebro humano aprende de uma forma mais lenta, criando ligações e pontes entre conjuntos de neurónios (Figura 7) com um determinado potencial medido em volts e formando estas ligações ao longo do tempo, começando com uma ligação fraca com várias opções, até obter a ligação definitiva. Este tipo de

aprendizagem permite ao cérebro humano ser flexível e incrivelmente poderoso a interpretar informação nova e descobrir novas maneiras de a utilizar.

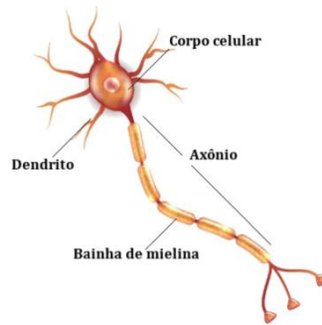


Figura 7 – Um neurónio cerebral humano (O que é o neurónio?, no date).

As redes neuronais criadas nos computadores actuais continuam a ser interpretadas pelos mesmos como instruções e activações de portas lógicas, sendo apenas programadas de uma maneira diferente da tradicional. A este tipo de redes neuronais chamamos redes neuronais artificiais (ANN's).

Esta ideia de imitar o cérebro humano surgiu em 1944 por Warren McCulloch e Walter Pitts, professores do MIT. Foi desenvolvida até 1969 com diferentes modelos, mas caiu em desuso até aos avanços recentes no poder de processamento das placas gráficas (GPU's) que voltaram a abrir o campo de forma exponencial.

O perceptrão foi o modelo mais simples de um classificador linear desenvolvido em 1959 por Frank Rosenblatt. Pode simplesmente ser descrito pela equação:

$$Valor\ de\ Saida = \begin{cases} 0 & se\ \sum_j w_j x_j + b \leq 0 \\ 1 & se\ \sum_j w_j x_j + b > 0 \end{cases} \quad (9)$$

nesta equação  $w_j$  é um factor de ponderação aplicado a cada uma das variáveis de entrada  $x_j$ .

Todos estes inputs são somados e passados a uma função de activação como a descrita na equação 9. O  $b$  é um factor adimensional que pode ser interpretado como a facilidade de fazer com que o perceptrão dispare. O perceptrão pode ser representado da seguinte forma (Figura 8):

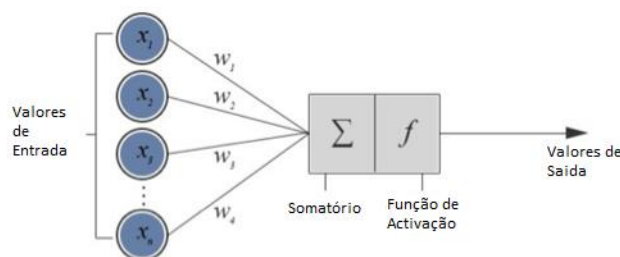


Figura 8 – Modelo de um neurónio artificial (Pokharna, 2016).

Uma rede neuronal pode ser constituída por perceptrões de várias camadas sendo então denominado por ANN (Pokharna, 2016).

### 2.5.1.2 – Elementos e parâmetros de uma ANN

Uma ANN é constituída por uma camada de entrada, uma ou várias camadas escondidas onde se processam os cálculos que convertem o input no output e uma camada de saída com o resultado (Kriesel, 2014). Todas estas camadas são constituídas por um ou mais neurónios dependendo da escolha do programador para cada situação. Na figura 9 apresenta-se um exemplo de uma rede neuronal simples com duas camadas escondidas.

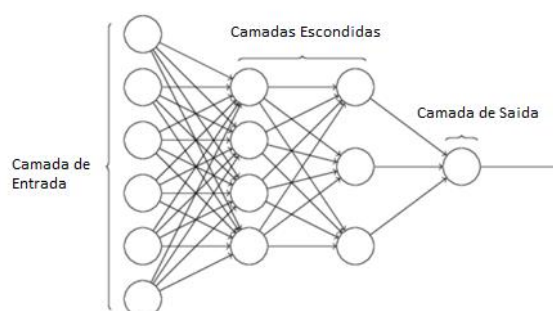


Figura 9 – Modelo de uma rede neuronal artificial (How Neural Networks Work, 2017).

Cada neurónio é um perceptrão, sendo o seu valor final passado a input ou combinação de inputs da camada seguinte e assim sucessivamente até chegar à camada final de saída. A este tipo de ANN dá-se o nome de “FeedForward” ANN, pois os valores circulam apenas num sentido e são as redes mais comuns (Goodfellow *et al.*, 2016). Existem outros tipos de redes, como as redes recorrentes, ou as convolucionais. Dizemos que uma rede neuronal é profunda quando tem mais do que uma camada escondida, podendo existir redes com mais de 100 camadas escondidas e milhares de neurónios em cada camada.

Para que uma ANN aprenda a resolver um problema precisa de adquirir experiência através dos dados de treino. Estes dados, já discutidos no capítulo anterior, devem ser pré-processados e classificados quanto ao valor esperado para servirem de métrica de controlo e correcção ao longo do processo de treino. A este processo de correcção, que afectará o valor dos pesos de cada neurónio por todas as camadas dá-se o nome de retro-propagação, e nele é calculado o erro do resultado da camada de saída face ao resultado previsto para aquele conjunto de dados. Esse erro é depois propagado proporcionalmente por todas as camadas anteriores até chegar ao input (Goodfellow *et al.*, 2016).

Portanto, basicamente o processo de treino envolve as seguintes etapas por cada linha de dados:

- uma amostra de treino é introduzida na rede e inicializada com pesos aleatórios;
- é calculado o output através da activação de neurónios ao longo das camadas escondidas até à camada final;



- após cálculo do erro, os pesos são reajustados por um algoritmo de retro-propagação cujo objectivo é reduzir o erro obtido anteriormente;
- o processo continua até os pesos estabilizarem ou até o conjunto de dados de treino ter terminado.

Um dos aspectos fulcrais em redes neuronais é o processo de retro-propagação, pois é este processo que é responsável pelas correcções que permitem à rede aprender até o seu nível de erro ser aceitável (Hopfield, 1988). O objectivo desta técnica é modificar os pesos de forma a minimizar o erro da função seleccionada para avaliar o erro de uma iteração ou conjunto de iterações da passagem “feedforward” pela rede. Existem vários algoritmos que podem proceder a esta correcção e podemos seleccionar vários até encontrar o adequado à rede em desenvolvimento. O conceito mais amplamente utilizado para algoritmos de retro-propagação é o da diferenciação. A diferenciação, aplicada à equação de erro permite obter a derivada da função, que por sua vez pode ser interpretada como a taxa a que a função está a alterar os seus valores num ponto específico (Kriesel, 2014; Pokharna, 2016).

O que nos interessa sempre é a taxa a que o erro muda, consoante as alterações aplicadas aos pesos da rede. Utilizando a abordagem matemática conseguimos obter esta informação muito mais rapidamente.

Um exemplo para um peso seria:

- calcular a derivada da função de erro;
- se a taxa for positiva para aquele valor de  $w$ , significa que o erro aumenta se aumentarmos os pesos, devemos reduzir o seu valor, e vice-versa;
- se a taxa for 0, atingimos o ponto de estabilidade máxima.

A figura 10 ilustra este processo.

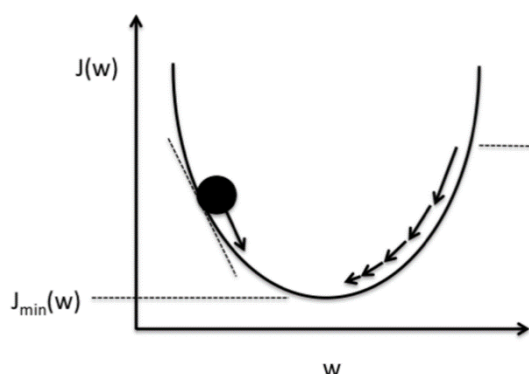


Figura 10 – Procura do mínimo absoluto da função de custo (Synced, 2017).

No caso de ANN's com várias camadas escondidas é necessário decompor a derivada da função de erro, começando na camada mais próxima até ao início (Abraham, 2015). Este

processo pode tornar-se computacionalmente muito intensivo para redes complexas, mas se conhecermos as funções aplicadas e as suas derivadas para cada camada, podemos corrigir os pesos para todas as camadas da ANN aplicando alterações em todos eles por diferentes métodos. Na figura 11 está representada a metodologia de treino de um modelo ANN.

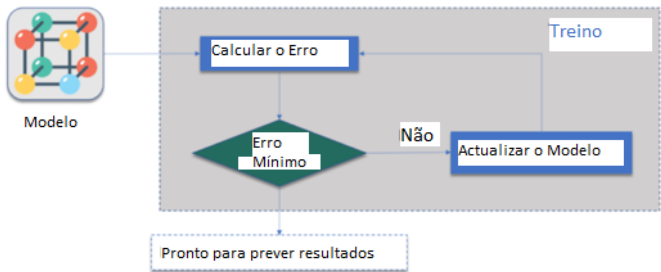
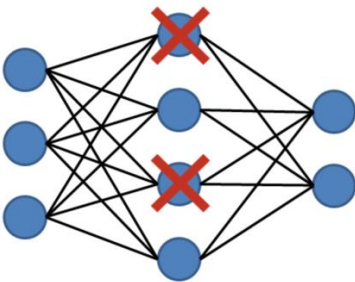
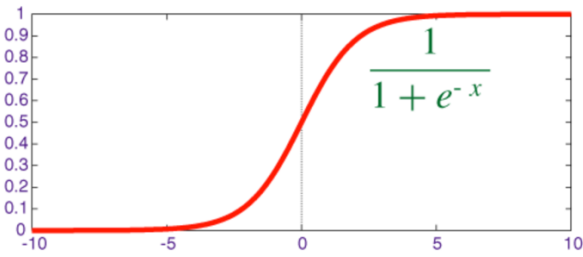
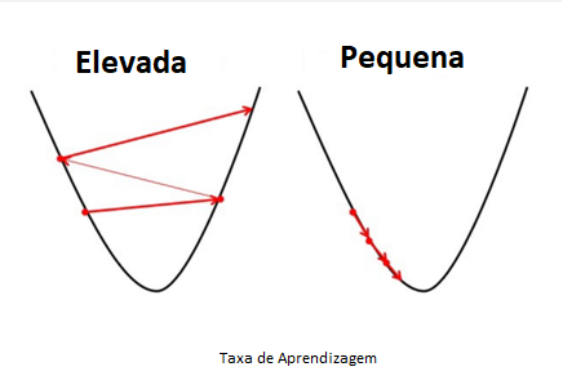


Figura 11 – Treino de uma ANN.

Outro aspecto igualmente importante prende-se com a parametrização de certas variáveis do modelo e do processo de treino que condicionam o comportamento e estrutura da rede. Estes parâmetros são designados por hiperparâmetros e são seleccionados antes de qualquer processo de treino(Zaccone *et al.*, 2018). Os principais são apresentados no quadro 5.

Quadro 5 - Parametrizações de uma Rede Neuronal(Goodfellow et al., 2016; Skansi, 2018)

Hiperparâmetro	Detalhes
Número de unidades e camadas escondidas	<ul style="list-style-type: none"> <li>- A selecção do número de camadas escondidas e unidades por camada.</li> <li>- Pode contribuir para subestimações se a rede for composta por poucas camadas</li> </ul>
“Dropout”	<ul style="list-style-type: none"> <li>- É uma técnica que permite regularizar a rede de forma a evitar a sobreestimação.</li> <li>- Permite remover alguns neurónios de serem utilizados de forma aleatória ao longo do treino, escolhendo apenas a percentagem de neurónios que se quer remover.</li> <li>- Especialmente eficaz em redes grandes.</li> </ul> 
Inicialização do valor dos pesos	<ul style="list-style-type: none"> <li>- Na maioria dos casos é utilizada uma distribuição normal para inicializar os pesos.</li> </ul>

<b>Função de activação</b>	<ul style="list-style-type: none"> <li>- Permitem introduzir modelos não lineares.</li> <li>- Existem inúmeras funções de activação de neurónios.</li> <li>- Requer avaliação prévia do problema a resolver para a sua selecção</li> <li>- Exemplo:</li> </ul>  <p style="text-align: center;">Sigmoid activation function</p>
<b>Taxa de aprendizagem</b>	<ul style="list-style-type: none"> <li>- Define a rapidez do processo de aprendizagem</li> <li>- Valores mais baixos aumentam o tempo de treino, mas permitem uma convergência suave dos valores</li> <li>- Valores maiores aumentam a velocidade do treino, mas podem conduzir a falta de convergência.</li> <li>- A preferência é por uma taxa de aprendizagem decrescente com o número de iterações.</li> </ul>  <p style="text-align: center;">Taxa de Aprendizagem</p>
<b>Momento</b>	<ul style="list-style-type: none"> <li>- Ajuda a conhecer a direcção do próximo passo, a partir do conhecimento adquirido sobre os anteriores.</li> <li>- Ajuda também a prevenir oscilações</li> <li>- Uma escolha típica é um valor entre 0.5 e 0.9</li> </ul>
<b>Número de “epochs”</b>	<ul style="list-style-type: none"> <li>- O número de vezes que o conjunto completo de treino deve ser passado pela rede durante o processo</li> </ul>
<b>Tamanho da amostra</b>	<ul style="list-style-type: none"> <li>- Define o número de amostras mais pequenas do conjunto de treino a considerar para que ocorra uma nova validação e alteração dos parâmetros da rede.</li> </ul>

Sumarizando, a figura 12 apresenta um resumo do processo de treino detalhado de uma rede neuronal.

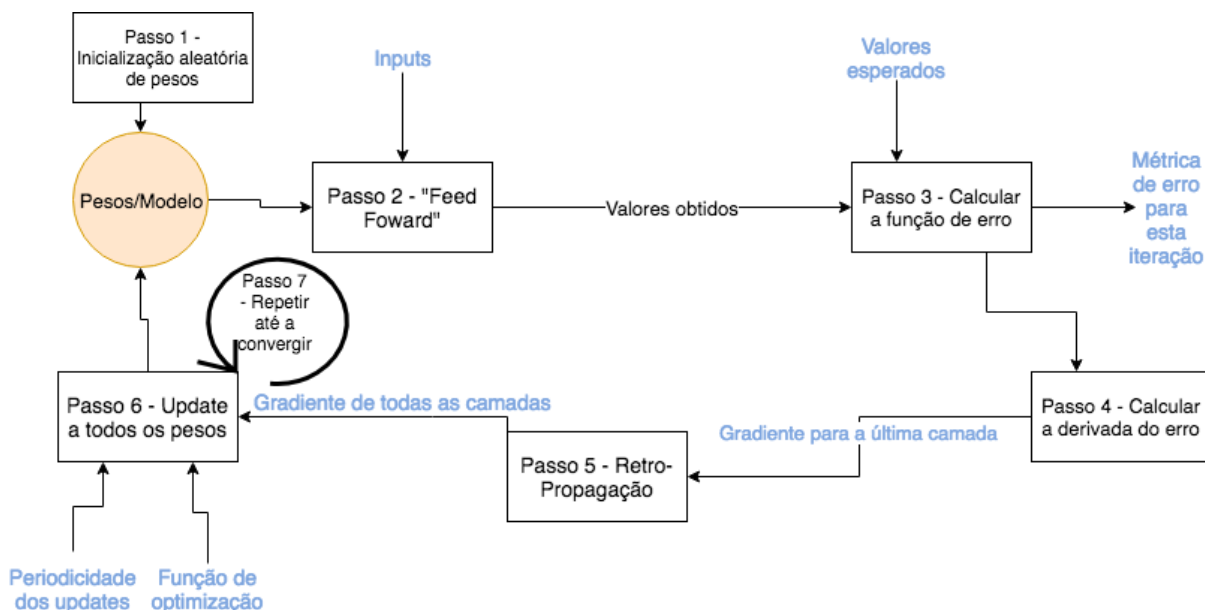


Figura 12 – Processo de treino detalhado de uma ANN.

## 2.5.2 – Aprendizagem reforçada

### 2.5.2.1 – Conceitos gerais

A aprendizagem reforçada (AR) tem suscitado grande interesse e avanços nos últimos anos, especificamente devido à proliferação de carros autónomos e novos algoritmos para os tornar mais seguros.

Esta abordagem combina a programação dinâmica, que é um campo da matemática tradicionalmente utilizado para a resolução de problemas de controlo e optimização e que se encontra limitado pelo tamanho e complexidade dos problemas que consegue resolver, e a aprendizagem supervisionada, onde é necessário conhecer as respostas certas ao problema para chegar a um modelo válido. Existem muitas situações em que é impossível saber estas respostas "a priori" e é aqui que entra a aprendizagem reforçada.

Não se pode definir a AR como sendo um tipo de ANN ou uma alternativa às ANN's, mas sim como uma ferramenta que pretende abordar problemáticas diferentes e mais difíceis de resolver. Ao combinar as duas abordagens referidas acima é possível construir sistemas poderosos e extremamente dinâmicos que permitem, através da definição de um objectivo, fazer com que o algoritmo aprenda a atingir aquele objectivo por tentativa e erro, interagindo com o ambiente que lhe foi fornecido, em vez de ter acesso a dados de treino classificados (Sutton *et al.*, 1998).

Por exemplo, consideremos a tarefa de ensinar um modelo computacional a andar de bicicleta. O seu objectivo é apenas andar de bicicleta sem cair. Para o atingir, o sistema vai executando ações como virar o volante à esquerda ou à direita consoante a posição em que se encontra. Ao tomar uma ação, o seu resultado é visível no estado da bicicleta, se continua de pé ou caiu, por exemplo. Consoante a resposta, o sistema toma outra ação, e assim sucessivamente, sendo recompensado quando toma a ação certa e sofrendo penalizações sempre que o ciclista cai e o objectivo foi falhado. Após milhares ou milhões de iterações, a bicicleta irá invariavelmente manter-se estável e será possível corrigir ações erradas, pois o modelo saberá reagir às mudanças (Saito *et al.*, 2018). Este tem sido o caminho utilizado por muitos investigadores, como os da *Google DeepMind* para construir uma inteligência artificial, uma vez que esta abordagem é a mais parecida com a abordagem humana a qualquer problema (Durugkar *et al.*, 2016; Mnih *et al.*, 2016). A grande vantagem do computador é conseguir, por simulações, aprender a uma taxa exponencialmente superior à dos humanos e atingir a proficiência em várias actividades diárias como já foi constatado pela mesma equipa da *Google* ao disputar o jogo Go contra o campeão da modalidade (Silver *et al.*, 2017).

### 2.5.2.2 – Formulação de um ambiente de AR

Num modelo de AR tradicional, um agente interage com um ambiente. Nesta interacção, o agente vai avaliando o ambiente e escolhendo uma ação a aplicar sobre o mesmo. A ação por sua vez altera o ambiente, sendo esta alteração comunicada de novo ao agente na forma de um sinal de AR. Logo existem três partes fundamentais de um sistema de AR: o ambiente, o agente que navega no ambiente e aprende e os estados em que o ambiente se encontra, com se segue.

- O Ambiente: define o conjunto de variáveis observáveis em determinada situação, cujas ações exercidas sobre essas mesmas variáveis lhe alteram o seu estado observável: exemplo: tabuleiro de xadrez, ambiente de jogos de computador.
- Agente: escolhe e executa ações no ambiente, recebe a informação sobre a alteração do estado do ambiente e das funções de reforço e recompensa.
- Estados: são constituídos por valores que os caracterizam individualmente face ao ambiente; quando o ambiente apresenta uma determinada combinação de valores, podemos codificá-los num estado, tornando a sua representação e interpretação mais concisa.

O objectivo principal de um sistema de AR é maximizar a recompensa final, ou seja, atingir o objectivo definido “a priori” com o máximo número de recompensas acumuladas em cada passo e ação, evitando dessa forma as ações com resultados negativos no ambiente e obtendo a solução ótima para todo o problema, sem qualquer ajuda (Sutton *et al.*, 1998).

Um dos desafios iniciais na criação destes sistemas é encontrar o balanço certo entre exploração/maximização. Se o agente estiver muito balanceado para maximização, irá ter

tendência a executar sempre as acções que lhe forneçam uma maior recompensa, mas podendo não atingir o resultado pretendido pois não explora o restante ambiente. Por esse motivo, é necessário incutir-lhe algum tipo de controlo aleatório, existindo várias estratégias para o conseguir. A estratégia “*epsilon-greedy*,” por exemplo, permite ao agente seleccionar uma acção ao acaso em vez daquela que seria mais provável de lhe maximizar a recompensa em 20% dos casos (Dutta, 2018). Normalmente esta percentagem é alterada ao longo do processo de treino, pois o agente vai aprendendo o caminho certo e reduzir o nível de acções aleatórias torna-se benéfico. É importante realçar que a recompensa pode ser só atingida após um longo conjunto de acções serem executadas. Na figura 13 encontra-se o diagrama simples de qualquer sistema de AR (Saito *et al.*, 2018).

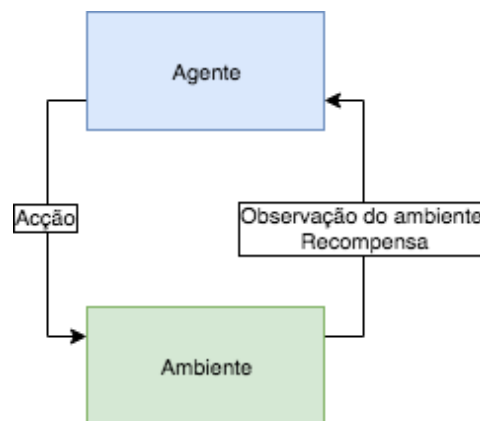


Figura 13 – Modelo de aprendizagem reforçada

Podemos definir formalmente as relações entre as acções, os estados e o agente como um *Processo de Decisão de Markov* ou MDP. Estes processos são definidos por:

- um conjunto finito de estados possíveis,
- um conjunto de acções por estado,
- transições entre estados,
- recompensas por cada transição,
- um factor de desconto  $\gamma$  entre zero e um, que quantifica a diferença entre a importância de recompensas imediatas ou futuras.

A função que traduz o objectivo de maximizar as recompensas ao longo do tempo pode ser dada por (Equação 10):

$$\sum_{t=0}^{\infty} \gamma^t r(x(t), a(t)) \quad (10)$$

em que  $\gamma$  é o factor de desconto,  $r(x,a)$  explicita a recompensa obtida para que no estado  $x$  se tome a acção  $a$ .

Formalizado o objectivo, o capítulo seguinte aborda uma possível solução para obter os resultados.

### 2.5.2.3 – Q-Learning

O *Q-Learning* é uma técnica que avalia qual a acção que deve ser tomada baseada numa função acção-valor que determina o valor de estar num certo estado e tomar uma determinada acção nesse mesmo estado. Existe uma função  $Q$  que recebe como *input* um estado e uma acção e devolve a recompensa esperada para aquela acção naquele estado. Antes de começar o processo de treino no ambiente, a função  $Q$  devolve sempre o mesmo valor fixo. Ao explorar o ambiente a função vai sendo actualizada até fornecer uma boa aproximação do valor da acção a no estado  $s$ . Cada estado tem a sua função  $Q$  que por sua vez é actualizada sempre que existe uma nova acção nesse estado (Watkins *et al.*, 1992). A função  $Q$  pode ser descrita da seguinte forma (Equação 11):

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha (r_t + \gamma \times \max_a Q(s_{t+1}, a) - Q(s_t, a_t)) \quad (11)$$

esta função implica que cada vez que tomamos uma acção no estado  $s$ , adicionamos ao seu valor um termo, que por sua vez contém os seguintes elementos:

- $\alpha$ : a taxa de aprendizagem, tal como nas ANN's (o conceito aqui é o mesmo): quanto maior o valor maior é o salto, neste caso do valor do par acção-estado;
- a recompensa,  $r_t$ , obtida por tomar a acção no estado;
- a estimativa da recompensa futura, que é dada pela recompensa máxima que se pode obter a partir de todas as acções disponíveis no estado para o qual a acção actual irá fazer transitar o agente;
- o valor actual de  $Q$  é subtraído neste termo, garantido que apenas se incrementa ou decrementa com base nas estimativas multiplicadas por  $\alpha$ .

Com base nos valores de  $Q(s_t, a_t)$  para todos os estados, o agente vai ajustando a sua estratégia até que todos os valores estejam otimizados e seja atingida a recompensa final da forma mais eficiente e com maior valor acumulado, garantindo que são tomadas as decisões que minimizem o número total de acções (Zaccone *et al.*, 2018).

Existem outras estratégias de resolução de MDP's, como a aprendizagem de políticas, em que se aprende uma função  $\pi$ , que mapeia directamente cada estado à melhor acção para aquele estado (Saito *et al.*, 2018).

Mais recentemente apareceram as "*Deep Q-Networks*" (DQN), que envolvem redes neuronais e *Q-Learning*, ou o algoritmo A3C, que combina a utilização de políticas e *Q-Learning* e representa o estado da arte na implementação de AR (Hausknecht *et al.*, 2015; Wang *et al.*, 2015).

## 2.6 – A oliveira (*Olea europeae* L.)

Estima-se que a oliveira seja cultivada desde 3000 A.C. um pouco por todo o mediterrâneo, e que ao longo dos séculos, os diversos povos que habitaram esta região foram descobrindo inúmeras utilizações para a azeitona e o azeite. Podem-se encontrar oliveiras em cerca de 98% da bacia do mediterrâneo, o que por si só é indicador da sua importância e proliferação na região, sendo hoje em dia uma cultura que se estende um pouco por todo o mundo onde seja possível encontrar um clima mediterrânico ou semi-árido (Giourga *et al.*, 2003).

A produção mundial de azeite tem vindo a crescer desde a década de 1990, onde se situava em 1335 milhares de toneladas. Actualmente e para o ano 2017/2018, a contagem provisória situa-se em 3133,5 milhares de toneladas, o que representa um crescimento substancial e uma aposta cada vez maior nesta cultura (dados obtidos de International Olive Council 2018).

Portugal situa-se no top 10 dos países produtores de azeite. A construção da barragem do Alqueva veio dar condições para que essa região do país se tornasse hoje em dia impulsionadora da produção nacional.

A balança comercial das exportações/importações é positiva desde 2010, contrariamente ao que acontece com a maior parte dos produtos agrícolas, em que não somos autossustentáveis (fonte: GPP).

A oliveira é tradicionalmente caracterizada pela sua rusticidade, suportando climas sub-áridos e solos muito pobres, apesar de estas condições terem um impacto negativo no potencial produtivo. O pH óptimo para o seu desenvolvimento é entre 6,5 e 8,5 e, se cultivada de modo tradicional, pode atingir uma grande longevidade (Oliveira, 2016). É, porém, sensível ao encharcamento e ao fraco arejamento do solo. Há cerca de 20/25 anos atrás, o olival nacional era sobretudo cultivado de forma tradicional e intensiva, neste caso com cerca de 200-300 árvores/ha, sendo as variedades tradicionais as mais utilizadas.

Hoje, a densidade mais utilizada é na forma de olival superintensivo com cerca de 900-2000 árvores/ha tendo sido introduzidas variedades adaptadas a este sistema como a “arbequina” e “arbošana”. Esta alteração, juntamente com a introdução da rega regular localizada, fertilização e uma actualização das técnicas de cultivo e colheita mecânica, tornaram o olival uma cultura altamente produtiva e rentável para os agricultores, com menor longevidade, mas atingindo a maturidade mais rapidamente.

Devido ao clima mediterrânico onde foi evoluindo ao longo dos milénios, a oliveira tem uma elevada tolerância ao stress hídrico, bem como ao frio e humidade invernal, e apresenta algumas características morfológicas e fisiológicas que a auxiliam no controlo do stress, permitindo-lhe taxas de transpiração moderadas. A diferença de potencial entre as folhas e a raiz pode ser de tal ordem que permite à planta extrair água do solo mesmo em condições que seriam limitantes para outras espécies. O fotossistema apresenta mecanismos de protecção que lhe permitem também manter a actividade fotossintética (Taylor, 2000).



Na figura 14 apresenta-se o ciclo anual vegetativo e reprodutivo da oliveira, sendo de realçar que uma das suas características são os anos de safra e contra-safra, o que a torna uma espécie com tendência a alternar a predominância dos ciclos (Sanz-Cortés *et al.*, 2002).

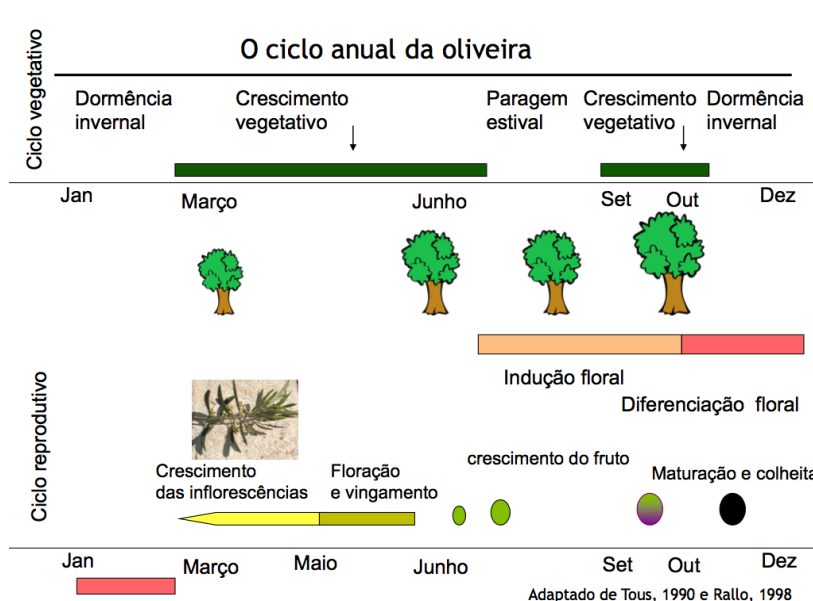


Figura 14 – Ciclo Anual Vegetativo da Oliveira. Adaptado de Tous, 1990 e Rallo, 1998

O fruto, a azeitona, é uma drupa e como tal o seu desenvolvimento após formação passa por 3 fases, uma fase I de multiplicação e expansão celular, uma fase II de endurecimento do caroço onde é possível a aplicação de stress hídrico para potenciar a qualidade do fruto, e uma fase III de crescimento rápido onde se dá também a biossíntese do azeite (Oliveira, 2016). Todo este ciclo dura cerca de 200 dias.

Até recentemente a oliveira sempre foi uma cultura de sequeiro mesmo tendo sido observado que a aplicação de pequenos volumes de água aumentava a produção substancialmente (Orgaz and Ferreres, 2007). A introdução da rega gota-a-gota transformou os olivais por todo o mundo, sendo que muitos dos olivais tradicionais foram convertidos para olivais modernos. Em 1990, 10% dos olivais era regado em Espanha. Em 2011 esse número já havia subido para mais de 25%.

Devido às alterações climáticas e à escassez de água referenciadas no capítulo 2.1, fornecer água suficiente para esta revolução no olival é um desafio e conduziu à proliferação de estratégias de rega deficitária, até porque não se prova que a rega de conforto traga benefícios. Foram realizados estudos sobre a adaptabilidade desta metodologia de rega às diferentes variedades, como a "arbequina" (Cuevas *et al.*, 2010; Palese *et al.*, 2010). Estas investigações respondem a um crescente interesse nas respostas da oliveira ao stress, pois a produção do ano seguinte depende do crescimento vegetativo do ano anterior que por sua vez, como na grande parte das árvores de fruto, é afectado pelo stress hídrico.

Por estas razões, e pela importância da cultura no panorama nacional é importante conseguir medir e interpretar as reacções da planta aos diferentes níveis de stress e aprender

como prevenir ou prever possíveis reacções adversas que conduzam a uma diminuição da produtividade e vigor.

Vários estudos têm sido desenvolvidos em Portugal sobre o uso da água em oliveira e indicadores de stress hídrico como por exemplo, (Santos *et al.*, 2007, 2013; Nogueira *et al.*, 2012) de que se destacam os realizados por equipas do Instituto Superior de Agronomia, por serem os únicos a incluírem medições de  $ET_a$ , em olivais regados e de sequeiro, por métodos de referência, como o das flutuações turbulentas (N Conceição *et al.*, 2017) e incluírem registos de indicadores do estado hídrico da planta (e.g. Ferreira *et al.*, 2012), bem como estudos sobre dinâmica da água no tronco e especialmente nas raízes através de sensores bidirecionais para avaliação e quantificação da redistribuição hidráulica, mecanismo de importância estratégica para a sobrevivência das plantas (Nadezhdina *et al.*, 2010, 2015, Ferreira *et al.*, 2012, 2018).

### 3 – MATERIAIS E MÉTODOS

#### 3.1 – Características do local

Os dados utilizados na elaboração deste estudo foram recolhidos na Herdade da Azambuja, Monte do Trigo (38° 24'N, 7° 43'W, 143 m, Figura 15).



Figura 15 - Delimitação da parcela em estudo.

O clima apresenta características de sub-húmido seco, correspondendo à classificação Csa, segundo o IPMA, que o caracteriza como clima temperado de tipo mediterrânico com verão muito quente e seco. A média anual de precipitação é entre os 600-800 mm, sendo a sua distribuição muito concentrada no tempo. A temperatura média mensal varia entre os 9.6 e os

24.1 °C. Os dados de temperatura (°C), precipitação (P) e  $ET_0$  no período de 2015 a 2018 podem ser visualizados na figura 16.

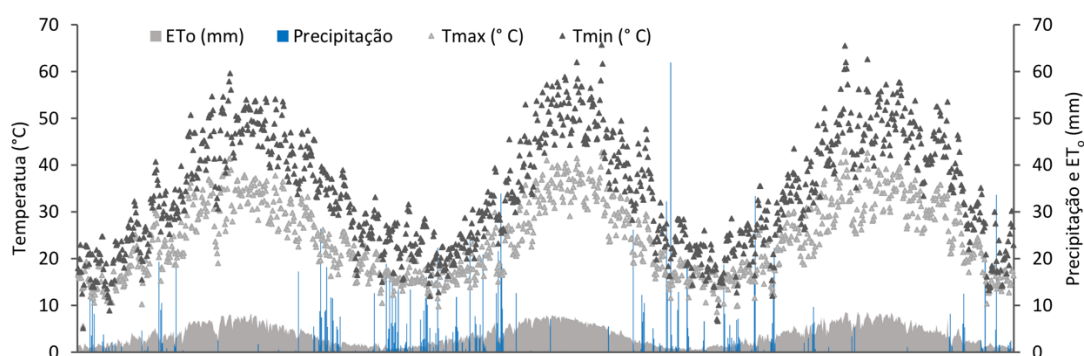


Figura 16 – Distribuição da precipitação, temperatura máxima e mínima e  $ET_0$  entre 2015 e 2018.

De acordo com a granulometria do solo (argila- 17%, areia fina-40%, areia grossa- 37%, limo-6%) cuja análise foi realizada em 2015, este é franco-argiloso.

O sistema de monitorização foi instalado em 2015 num olival superintensivo da variedade “Arbequina”, plantado em 2006. A plantação tem as linhas orientadas na direcção Sul-Norte, o talhão em estudo ocupa 78 ha e o compasso utilizado é de 1.35 x 3.75 m (1975 árvores/ha).

O sistema de rega gota-a-gota é composto por uma rampa por linha e espaçamento entre gotejadores de 0.75 m. Os gotejadores não são autocompensantes e apresentam um caudal teórico de 2.3 L/h, com um raio molhado ( $f_w$ ) de 0.23 m. A pressão de funcionamento dos mesmos é 0.1 MPa (1 bar). A rega é aplicada com uma frequência elevada (diária), de março a setembro, podendo ser ajustada consoante as condições microclimáticas. A água utilizada é proveniente de um furo de captação dentro da parcela em estudo, sendo o sistema delimitado por 10 unidades de rega activadas por electroválvulas controladas a partir de um programador central. Em condições normais a rega é efectuada de forma sequencial entre os sectores.

## 3.2 – Instrumentos de medição e aquisição de dados

### 3.2.1 – Sistemas de aquisição de dados

Os sistemas de aquisição de dados (SAD, vulgo *logger*) utilizados seja na estação meteorológica ou ligados a sensores instalados na parcela experimental suportam até oito entradas de sinal, três digitais e cinco analógicas (Modelo: ML-2013, Marca: YDOC, Origem: Holanda). A comunicação é feita por um modem de GSM 4G instalado na placa-mãe. O seu software permite configurar o formato de envio dos dados para vários pontos em simultâneo (*HTTP, FTP, MQTT*) e a execução de algumas funções matemáticas básicas sob os inputs sensoriais.

Os SAD estão configurados para recolher novos dados de 15 em 15 minutos e transmiti-los de 3 em 3 horas. Em caso de falha de comunicação, a mesma é armazenada num cartão SD pré-instalado no SAD até a comunicação ser reposta; são alimentados por pilhas alcalinas AA,

recarregáveis por um painel solar instalado na caixa impermeável que envolve os componentes. O seu tempo útil de utilização em condições normais é superior a 2 anos (Figura 17).



Figura 17 – SAD da estação de medição.

### 3.2.2- Sonda de capacitância Enviropro® (80 cm)

A sonda utilizada para medir a humidade no solo é do modelo EP100GL-08 *Enviropro*® sendo constituída por um sensor de capacitância, um sensor de salinidade e outro de temperatura a cada 10 cm do seu comprimento, num total de 80 cm. Os 8 sensores estão inseridos num bloco de cerâmica resistente ao choque e impermeável. A ligação ao SAD é feita por SDI-12, permitindo alimentação e transmissão de dados de forma contínua (figura 18).

A sonda foi instalada na região do bolbo molhado, a cerca de 10 cm do gotejador mais próximo e a 40 cm do tronco da oliveira.

Estas sondas possuem um mecanismo de calibrações internas que permitem compensar o teor de humidade medido com base nas variações de salinidade e temperatura do solo. O modelo vem pré-calibrado de fábrica. No quadro 6 encontram-se as especificações da sonda.

Quadro 6 –Especificações da Sonda Enviropro®

Leitura	Unidade de Medida	Resolução	Precisão
<b>Humidade</b>	%	0.01 %	± 2%
<b>Salinidade</b>	dS/m	0.001 dS/m	± 5 dS/m
<b>Temperatura</b>	°C	0.01 °C	± 1 °C

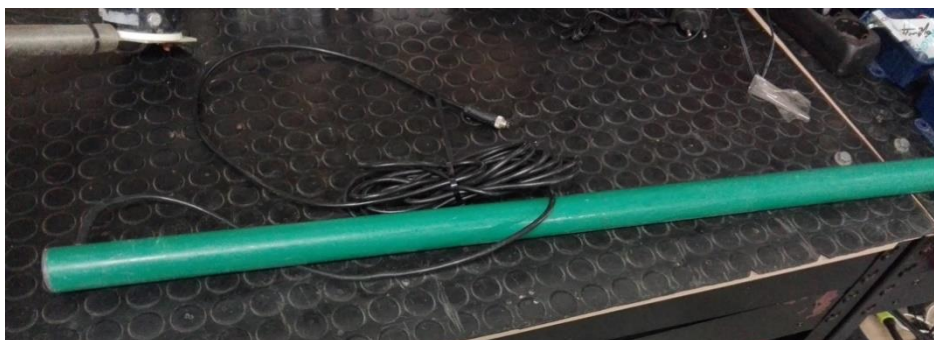


Figura 18 – Sonda Enviropro® antes da instalação.

### 3.2.3 – Caudalímetro

O caudalímetro utilizado para registrar os eventos de rega (Modelo: GSD8, Marca: B-Meter, Origem: Itália), tem como princípio de funcionamento a transformação das rotações analógicas para sinal digital através da emissão de pulsos magnéticos no seu interior, considerando-se um pulso como a passagem de um litro pelo tubo do ramal, e contando esses pulsos com um detector na superfície (figura 19).



Figura 19 – Caudalímetro B-meter.

### 3.2.4 – Dendrômetro

O dendrômetro de tronco digital, utilizado para registrar as variações do tronco da oliveira (Modelo: DE-1M, Marca: Bio Instruments S.R.L, Origem: Roménia) é altamente preciso e permite detectar variações na ordem dos microns (Figura 20). O sensor é montado num suporte de invar afixado ao tronco da árvore. O seu alcance permite medir variações no intervalo de 0 a 10 cm e a sua resolução é de 0.005 mm. Necessita de ser reajustado sempre que o crescimento atinge o limite de 10 cm. O sensor mede diferenças de potencial que se convertem em milímetros pela seguinte relação  $\Delta R = 5 \times U$ , onde  $U$  é a diferença de potencial em volts e  $\Delta R$  a variação no raio do tronco (micrómetros).





Figura 20 – Dendrômetro instalado no tronco de uma oliveira da estação de medição.

### 3.2.5 – Estação de medição

Considera-se uma estação de medição, o conjunto composto pelo SDA, sonda *Enviropro*, caudalímetro e dendrômetro.

Esta estação foi instalada numa linha da parcela descrita no subcapítulo 3.2.1. O SAD foi instalado num mastro elevado de forma a obter a melhor exposição solar no painel solar e a sonda foi instalada sob a rampa de rega a aproximadamente 40 centímetros do tronco. Para a instalação do dendrômetro, escolheu-se uma árvore sem a presença aparente de agentes patogénicos e de vigor mediano em relação ao restante olival.

### 3.3 – Pré-tratamento de dados

Os dados recolhidos por via *FTP*, são armazenados numa base de dados *MySQL* e encontram-se disponíveis para tratamento a partir desse momento.

Neste trabalho procedeu-se à recolha de todos os dados meteorológicos, sensoriais e de parametrizações da cultura, armazenados na Base de Dados (BD) em formato csv.

Os dados disponíveis abrangem o período de 16-05-2015 a 24-09-2018, o que inclui quatro ciclos da cultura, embora o primeiro e o último sejam incompletos. Para os dados do ano de 2017, a estação de medição esteve sem comunicações do dia 2017-11-02 até ao fim do ano, e como tal os dados analisados nesse ano apenas abrangem o período até essa data. A hipótese de interpolar os dados foi posta de parte, pois a criação de valores para aquela fase do ano não se justificava e iria introduzir uma fonte de erro adicional às análises em curso.

Procedeu-se de seguida à sua importação para *MS Excel* e à sua divisão em séries anuais.

As leituras de 15 em 15 minutos sofreram um processo de “*resampling*” uma vez que a gestão de rega é uma actividade diária e, para os ensaios a realizar, não é necessária uma periodicidade tão pequena. Para o efeito foram escolhidos o valor máximo de cada dia como a leitura a reter para cada variável obtida através dos sensores instalados e descritas no quadro 7.

Quadro 7 – Unidades das variáveis recolhidas no SAD.

Variável	Unidade
Diâmetro máximo diário do tronco (MaxD), diâmetro mínimo diário do tronco (MinD), registos de rega (RR), precipitação (P)	milímetros
Velocidade média do vento (Vmed), velocidade mínima do vento (Vmin), velocidade máxima do vento (Vmax)	m/s
Humidade relativa média, humidade relativa máxima, humidade relativa mínima	%
Temperatura média do ar (Tmed), máxima (Tmax) e mínima (Tmin)	°C
Somatório da humidade medida por todos os sensores individuais da sonda (HSUM)	% (V/V)
Temperatura do solo a diferentes profundidades (10 - 80 cm)	°C

Os dados recolhido no SAD foram utilizados para o cálculo de  $ET_0$  diário pelo software Irristrat®. Os valores de  $K_c$  simples (Quadro 8) utilizados foram também fornecidos pelo mesmo software, garantindo que o agricultor e o algoritmo inteligente a desenvolver (próximos capítulos) partilham a mesma base de informação na tomada de decisão.

Quadro 8 – Valores de  $K_c$  ao longo do ciclo cultural.

Fase	Nº de dias	$K_{c\ inicial} - K_{c\ final}$	Data de início
1	61	0.3-0.3	1 Janeiro
2	28	0.3-0.5	3 Março
3	31	0.5-0.5	31 Março
4	91	0.5-0.3	1 Maio
5	54	0.3-0.3	31 Julho
6	8	0.3-0.5	23 Setembro
7	61	0.5-0.5	1 Outubro
8	31	0.5-0.3	1 Dezembro

Durante o período de crescimento do tronco, foi necessário ir realizando ajustes ao dendrómetro. Estes ajustes provocam o aparecimento de saltos anormalmente altos nos valores medidos. Como tal, e como o interesse está na dinâmica destas variações, os valores foram normalizados com a seguinte equação:

$$Factor\ de\ Correção = DiamMax_d - DiamMax_{d-1} \quad (12)$$

em que  $DiamMax_d$  é o diâmetro máximo do tronco medido no dia do ajuste e  $DiamMax_{d-1}$  o valor máximo no dia anterior ao ajuste.

Estas variáveis foram posteriormente importadas para o ambiente de desenvolvimento e analisadas.

### 3.4 – Ambiente de desenvolvimento de algoritmos

A análise dos dados recolhidos, todo o seu tratamento e construção de algoritmos foi executada na linguagem de programação *Python*, criada em 1991 por Guido van Rossum. Esta é uma linguagem de programação “*High-level*” interpretada. A sua interpretação e leitura é bastante fácil do ponto de vista da sintaxe. É distribuída de forma gratuita pela *Python Software Foundation*.

No meio académico e científico, o *Python* é amplamente utilizado existindo bastantes bibliotecas avançadas, dedicadas ao *Big Data*, ML e construção de ANN’s. Neste trabalho foram utilizadas algumas destas bibliotecas apresentadas no quadro 9.

Quadro 9 – Bibliotecas e Frameworks utilizadas no trabalho.

Nome	Detalhes da utilização
Jupyter	-Visualização interactiva de dados e gráficos no browser
Pandas	- Construção, manipulação e análise de grande volume de dados em formato semelhante ao MS Excel
scikit -learn	- Modelos de Machine Learning
TensorFlow	- Construção de Redes Neurais

### 3.5 – Metodologia de análise dos dados de dendrometria

Um dos objectivos deste trabalho é a procura de um índice que permita que os valores medidos pelo dendrómetro funcionem como um bom indicador do stress hídrico da planta.

Para ser um bom indicador, este índice deverá ser facilmente interpretável e comportar-se de acordo com as dinâmicas observadas por outros indicadores como por exemplo as leituras da sonda, funcionando como um alerta precoce para situações de stress e da sua intensidade. A metodologia aqui proposta tem como propósito observar todos os inputs sensoriais, seleccionar os que mais influenciam os valores e índices derivados dos dados do dendrómetro e criar um algoritmo que seja eficiente a detectar e prever os níveis de stress a partir dos inputs disponíveis.

O processo de selecção do índice encontra-se sumarizado na figura 21.



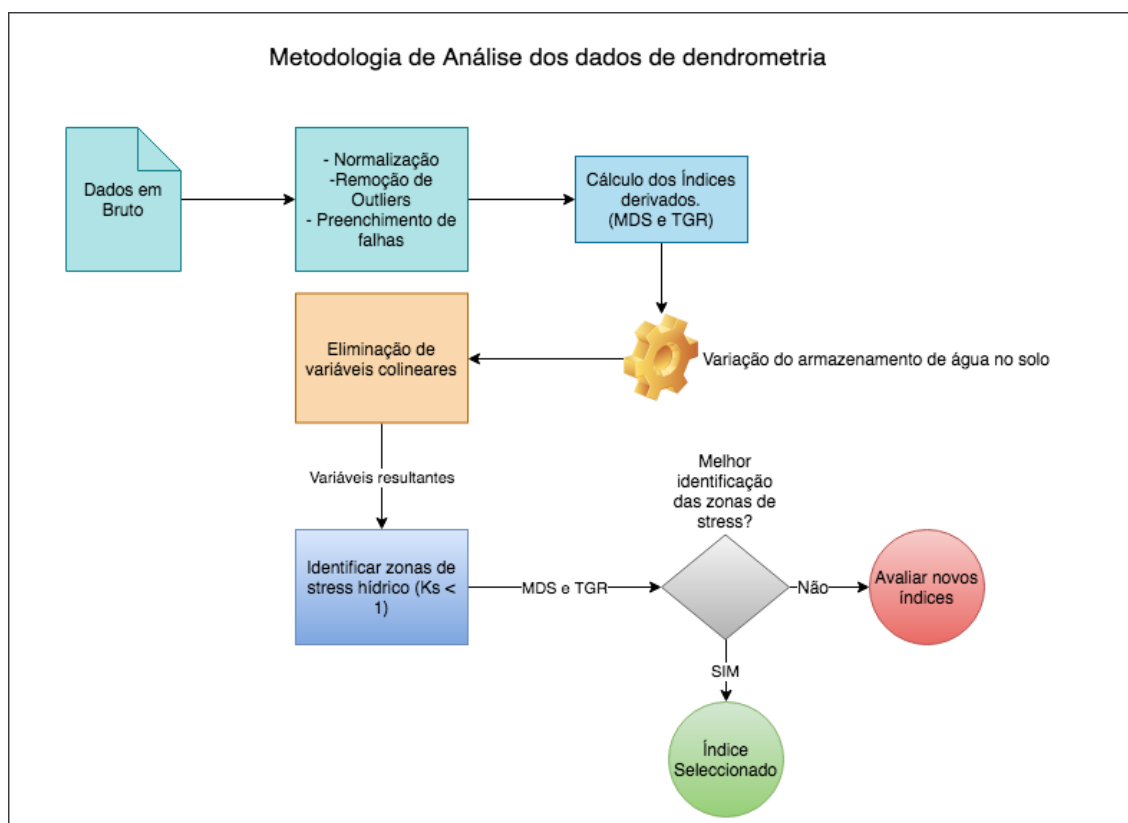


Figura 21 – Metodologia de análise da dendrometria.

Os dados dos 2 ciclos culturais completos disponíveis são separados e analisados individualmente pelos mesmos passos, descritos de seguida:

1. Cálculo dos Índices mais comuns em dendrometria segundo a literatura (MDS e TGR) a partir das medições.
2. Construção da variação de armazenamento de água no solo a partir do balanço hídrico diário e estimativa de  $K_s$ :

Os dados de parametrização para avaliação da reserva de água no solo foram os seguintes (Quadro 10):

Quadro 10 – Dados recolhidos no campo e estimados do projecto OlivoReg.

Variável	Valor	Unidade
$\Theta$ para capacidade de campo (CC)	0.24	$\text{cm}^3/\text{cm}^3$
$\Theta$ para coeficiente de emurchecimento (CE)	0.12	$\text{cm}^3/\text{cm}^3$
Profundidade radicular ( $Z_r$ )	0.8	Metros
Défice permissível ( $p$ ) base	0.4	Adimensional

A capacidade de campo e coeficiente de emurchecimento foram obtidos por análise gravimétrica realizada pelo ISA no âmbito do projecto OlivoReg, o valor de  $p$  de base foi também seleccionado a partir da experiência de rega em olival da empresa Hidrosoph.

A profundidade radicular foi fixada em 0.8 metros, pois as plantas em estudo são adultas, num sistema de plantação superintensivo onde a competição por recursos é elevada e os crescimentos radiculares são limitados pela presença das árvores vizinhas, logo admitiu-se que  $Z_r$  não sofra alterações ao longo das campanhas. A incerteza quanto ao valor de  $z$  constitui uma motivação para o uso de fito-sensores.

Calculou-se  $RU$ ,  $RFU$ ,  $p_{correctado}$ ,  $\Delta S$ ,  $R_d$ ,  $K_s$  e  $ET_a$  pelas expressões:

$$RU = (CC - CE) \times Z_r \times 1000$$

$$RFU = RU \times p$$

$$p_{correctado} = 0.4 + 0.04 \times (5 - ET_c)$$

$$\Delta S = P + R - ET_a \quad (13)$$

$$R_d = \sum ET_a$$

$$K_s = \begin{cases} 1, & \text{se } R_d < RFU \\ \frac{RU - R_d}{RU - RFU}, & \text{se } R_d > RFU \end{cases}$$

$$ET_a = K_s \times K_c \times ET_0$$

Para o  $\Delta S$  ignoram-se as perdas por escoamento superficial, drenagem e percolação para simplificar os cálculos.

3. Caracterização estatística de cada variável e a sua distribuição.
4. Matriz gráfica de relações entre as variáveis, eliminando as que não apresentam qualquer relação com os índices calculados.
5. Se presente, subdividir os dados para o período em que o valor de  $K_s$  determinado empiricamente é menor que 1, o que por sua vez estima a entrada da planta num período de stress hídrico. Verificar qual o comportamento dos índices existentes neste período e caracterizá-los.

6. Escolher o índice com maior sensibilidade aos períodos de stress e quantificar o seu valor nesses mesmos períodos.
7. Após esta análise nos dois anos completos de dados, calcula-se o valor médio do índice mais sensível, eliminando grande parte da variabilidade intrínseca às condições específicas de cada ano.

De seguida testa-se a viabilidade de utilizar dois modelos de ML, uma rede neuronal simples e uma floresta de decisão aleatória, para prever os valores de MDS expectáveis a partir de variáveis ambientais e do solo, que são as mais facilmente disponíveis. Essas variáveis serão,  $ET_0$ , MDS, HSUM, variação do armazenamento da reserva do solo e estados fenológicos.

O conjunto de dados de treino dos modelos é composto por 75% dos 4 anos de dados recolhidos.

A avaliação dos modelos é feita utilizando os restantes 25% do conjunto de treino e avaliando o RMSE da classificação.

### **3.5.1 – Floresta de Decisão aleatória**

Um algoritmo de ML deste tipo utiliza um conjunto de árvores de decisão simples criadas com partes aleatórias do conjunto de dados já classificados, sendo um algoritmo de aprendizagem supervisionada.

Uma árvore de decisão individual representa a avaliação de uma condição ou estado com base em variáveis e nos seus valores. Pode ser usada em problemas de regressão ou classificação. Cada nó representa a avaliação de um valor de uma ou mais variáveis, podendo conduzir a outro nó ou a vários nós diferentes. O primeiro nó representa a variável cujo valor é mais relevante para o resultado e assim sucessivamente. No fim do processo a árvore conduz a uma previsão.

O resultado final do algoritmo FDA é a previsão mais vezes obtida pelo conjunto das árvores individuais. Desta forma reduz-se o risco de sobrestimação dos resultados. Quanto maior for o número de árvores individuais, mais robusto e preciso é o resultado.

Para a construção deste modelo, pretendemos obter os valores de MDS a partir das variáveis referidas acima.

Para a implementação do algoritmo, os dados de treino são submetidos a algumas normalizações. Os valores da sonda são normalizados de 0 a 1, correspondendo 1 à capacidade de campo e 0 ao ponto de emurchecimento. Os restantes valores são mantidos nas unidades de obtenção.

O algoritmo final é constituído por 5000 árvores de decisão individuais, sendo cada uma delas responsável por uma previsão a partir das variáveis de input. Este número de árvores foi selecionado após experimentação de várias configurações.

### 3.5.2 – ANN

As redes neurais oferecem uma abordagem diferente aos algoritmos de classificação tradicionais, podendo, ou não, apresentar maior precisão do que estes. Para esta fase do trabalho e após implementação da FDA, foi criada uma rede neuronal simples com duas camadas. Todos os perceptrões da primeira camada estão ligados às variáveis de input que por sua vez, são idênticas às utilizadas para o treino da FDA. No entanto para uma melhor convergência no processo de treino desta rede, todos os valores foram reduzidos pela equação 14 reduzindo o efeito de valores “outlier”:

$$V_{reduzida} = \frac{V_{real} - \mu}{\sigma} \quad (14)$$

em que  $\mu$  é a média dos valores da variável e  $\sigma$  o seu desvio padrão, permitindo assim aos valores reduzidos assumir uma distribuição normal.

A função de activação de cada perceptrão é definida por camada. A escolha desta função deve ser feita de acordo com o tipo de dados e se estamos perante um problema de regressão ou classificação. Neste caso, para a primeira camada usa-se a função tangente hiperbólica definida por  $f(x) = \tanh x$ . Esta função permite normalizar o output de cada neurónio entre -1 e 1. Sendo que o input é constituído por valores positivos e negativos com muita amplitude entre os valores extremos, faz sentido normalizar esses valores num intervalo limitado que desta forma suaviza a influência dos extremos.

A segunda camada que está ligada ao output final da rede utiliza a função RELU (Unidade linear rectificada). Esta função é definida por  $R(z) = \max(0, z)$  e o principal propósito da sua utilização é impedir que todos os perceptrões sejam activados ao mesmo tempo. Para o propósito deste algoritmo esta função permite introduzir maior variabilidade no processo de treino, reduzindo o problema de sobreestimação. Hoje em dia é a função mais utilizada. Na figura 21 estão representas as duas funções de activação.

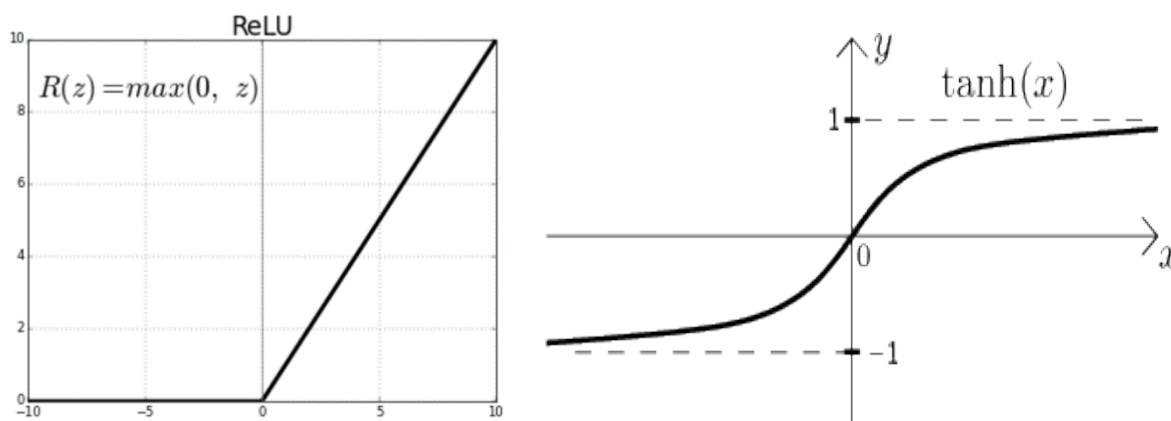


Figura 22 – Funções de Activação da ANN.

A função de custo escolhida foi o erro quadrático médio, avaliado a cada iteração. A função de optimização corrige o valor dos pesos procurando o mínimo absoluto da função de custo. Quanto mais rápido for a atingir o mínimo absoluto, mais eficiente é o algoritmo.

Como o volume de dados é reduzido (aproximadamente 1000 pontos) é necessário utilizar um algoritmo que restrinja o cálculo do gradiente em cada ponto. Para tal foi escolhido o *RMSprop* que restringe as oscilações do gradiente num dos planos. Se o espaço for 2D, imaginemos que a procura do mínimo absoluto se faz passo a passo, e que cada passo é maior em y do que x (Figura 23). Ao restringir a direcção do gradiente podemos garantir o objectivo de optimização é atingido mais facilmente.

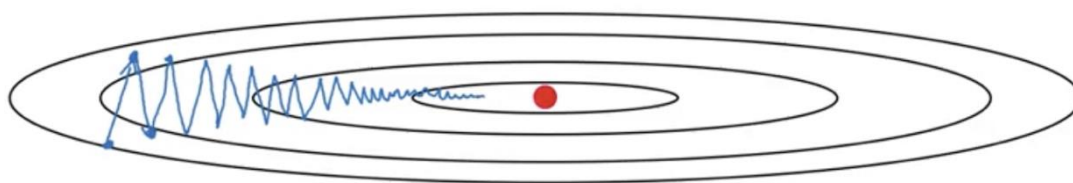


Figura 23 – Mecanismo de procura do mínimo absoluto da função em 2D.

As métricas de avaliação da performance serão o erro médio absoluto e a precisão, tal como na FDA.

### 3.6 – Metodologia de AR para gestão inteligente da rega

A aprendizagem reforçada oferece a possibilidade de criar modelos de ML que evoluam com o seu ambiente e que estejam em constante processo de aprendizagem, não ficando estagnados no momento do seu treino com um conjunto limitado de dados. O conceito de recompensa, seja a cada passo, ou apenas ao atingir o objectivo final, é facilmente adequado ao meio agrícola e especialmente à gestão de rega, pois existe sempre um desfaseamento entre as acções e as suas consequências no estado hídrico da planta.

Vejamos, um modelo de AR necessita de um ambiente, simulado ou real, cujas reacções a acções executadas sobre esse mesmo ambiente possam ser quantificadas de forma precisa. Ao visualizarmos o conjunto cultura, solo, clima e sistema de rega estamos perante um ambiente que podemos quantificar e definir razoavelmente bem. Podemos também definir as acções como estratégias de condução da rega e medir as alterações ao ambiente com base nas redes sensoriais instaladas no campo e nas plantas. Para as recompensas pode definir-se a manutenção da planta na zona de conforto hídrico ou outra qualquer variável, como o sucesso em evitar a entrada da planta em stress hídrico.

Existem várias formas de abordar a formulação de um algoritmo deste género. Neste trabalho apresentam-se um conjunto de pressupostos para a criação de um algoritmo versátil e robusto que possa ser implementado na condução diária da rega.

### 3.6.1 – Ambiente

Na definição do ambiente é importante considerar variáveis que descrevam um ou mais fenómenos do ecossistema cultural, tendo o cuidado de não selecionar variáveis com elevada correlação entre si. É necessário incluir elementos da dinâmica do solo, do clima e da planta como descrito inicialmente. As variáveis meteorológicas e os seus efeitos, por uma questão de simplicidade, podem ser traduzidas numa única variável,  $ET_0$ , eliminando assim a necessidade de considerar o efeito do vento, temperatura e humidade separadamente.

Para a dinâmica do solo podemos utilizar os dados obtidos a partir da sonda se estivermos em ambiente de testes no campo. Como estamos perante uma abordagem teórica com dados históricos, é necessário simular esta dinâmica através dos dados da variação do armazenamento, pois este pode ser ajustado consoante as acções escolhidas ao longo do tempo.

Quanto à dinâmica da planta, e tendo em conta o ensaio descrito no capítulo anterior, iremos utilizar o modelo de ML escolhido no mesmo para simular o estado hídrico da planta a partir das variáveis existentes.

O ambiente neste ensaio será então modelado por  $ET_0$ ,  $R_d$ ,  $\Delta S$ , pelo modelo de da planta e pelos meses do ano.

### 3.6.2 – Acções

As acções para um sistema deste género necessitam de ser adequadas ao problema e provocar alterações visíveis e quantificáveis no ambiente. Neste caso as acções serão eventos de rega, distribuídos numa série de valores, desde zero a sete milímetros, com 12 valores incrementais intercalares. A escolha de sete como máximo não foi arbitrária, tendo muito em conta os valores máximos de  $ET_0$ , o facto de a rega poder ser aplicada diariamente sem restrição e a cultura em causa ser resistente ao stress hídrico, beneficiando até, de um stress moderado em determinadas fases do seu ciclo (Palese *et al.*, 2010) .

### 3.6.3 – Estados

Ao executar uma acção, esta terá repercussões no ambiente e o estado em que o ambiente se encontra sofrerá alterações. Para este problema precisamos de definir quais as condições do ambiente que definem cada estado. Iremos definir estes estados como combinações de intervalos de valores das variáveis disponíveis. Estes estados encontram-se descritos no quadro 11.

Quadro 11 – Definição dos estados

Estado	ET0	$\Delta S$	Planta ML	Reserva(%v/%v)
CH	--	Indiferente	Não Stress	100%-80%
A1	--	Indiferente	Não Stress	80%-70%
A2	--	Indiferente	Não Stress	70%-60%
A3	--	Indiferente	Não Stress	60%-50%
PS	--	Indiferente	Stress	< 80%
R1	--	Negativo	Stress	< 70%
R2	--	Negativo	Stress	< 60%
R3	--	Negativo	Stress	< 50 %
Falha1	--	Indiferente	Indiferente	< 40%
Falha2	--	Negativo	Stress	< 40%

A nomenclatura de cada estado representa as diferentes situações, sendo que deve-se considerar o estado CH como conforto hídrico, os estados A1 a A3, como zonas onde a planta não se encontra limitada por qualquer um dos factores em consideração. O estado PS como indicador de stress obtido a partir da dendrometria. Os estados R1 a R3 como situações onde a planta é afectada negativamente pelos valores das variáveis analisadas. E finalmente, os estados Falha1 e Falha2 como as condições necessárias para que o algoritmo considere a execução falhada, ou seja, foram falhados os objectivos propostos para a condução da rega.

O conjunto destes estados traduz as várias situações em que é necessário ajustar as dotações de rega, podendo existir muitos mais estados e combinações em situações reais. Para as condições deste teste, as apresentadas são suficientes.

### 3.6.4 – Recompensas

Tal como foi descrito no início do capítulo, o sistema avalia as acções e os estados através da atribuição de recompensas. Ou seja, mudanças de estados negativos para positivos são recompensadas, podendo até estas recompensas serem ajustadas aos objectivos de gestão. Para o ensaio consideremos que transições para o estado A1, A2 ou CH recebem recompensa positiva. Transições para o estado A3 recebem penalização mínima, para o estado PS ou R3, penalização máxima e para o R1 e R2 penalização intermédia.

Quadro 12 – Valores de recompensa em cada estado

ESTADO	RECOMPENSA
CH	0
A1	125
A2	100
A3	50
PS	-50
R1	-100
R2	-125
R3	-150
F1	-500
F2	-500

Atingir qualquer um dos estados Falha, implica que o conjunto de acções escolhidas falhou, a simulação é parada e a recompensa é 0. O mapa de recompensas é apresentado no quadro 12.

### 3.6.5 – Estratégia de Q-Learning

Considerando as regras descritas acima criou-se uma simulação do ambiente em *python* utilizando todos os dados recolhidos do olival no período de 2015 a 2018. O sistema permite simular a evolução da reserva útil e da sua variação diária a cada iteração durante o treino, permitindo assim controlar o ambiente e o efeito das acções sobre o mesmo. Os dados são também introduzidos no modelo de ML de previsão do valor de MDS a cada iteração.

O algoritmo utiliza a técnica de AR denominada por DQN. Esta técnica utiliza uma rede neuronal para prever os valores Q em cada par Estado-Acção. Ao reproduzir milhares de vezes o ambiente, a rede neuronal aprende as melhores combinações a seguir e a avaliar correctamente qual a acção a tomar em cada estado. Para além disso esta técnica introduz o conceito de recompensas retardadas.

O sistema toma uma acção, neste caso, escolhe a quantidade de água para um evento de rega, recebendo em simultâneo a recompensa se mudou de estado hídrico e todas as recompensas futuras para todos os pares (estado,acção) possíveis de atingir a partir do novo estado, sendo estas ponderadas por um factor de desconto a definir.

Para garantir que o algoritmo converge para valores positivos de recompensa durante o treino são utilizadas duas técnicas, a repetição de simulações passadas e a utilização de duas redes neuronais iguais, mas actualizadas a intervalos de tempo diferentes.

A repetição de simulações passadas permite ao algoritmo criar a sua própria memória de iterações, guardando para isso todos os dados da mesma. Estes valores são armazenados e são introduzidos no processo de treino em pontos aleatórios, contribuindo para que a experiencia acumulada na resolução do problema aumente significativamente.

Em simultâneo, a utilização de duas redes neuronais para estimar os valores Q permite que enquanto uma rede é actualizada a cada iteração, a segunda rede aja como um acumulador, que só é actualizado após vários episódios completos.

Todas estas considerações e detalhes tornam o algoritmo implementado uma solução completa de aprendizagem reforçada.



## 4 – RESULTADOS

### 4.1 – Análise de dendrometria para detecção do stress hídrico

De acordo com o definido no capítulo anterior, procedeu-se à análise dos dados sensoriais recolhidos durante os vários anos. A análise começa pelo primeiro ano completo, 2016, depois 2017. Os dados relativos a 2015 e 2018 irão servir para validar os modelos de ML criados.

#### 4.1.1 – 2016

Para o ano 2016 as variáveis em análise apresentam as seguintes características estatísticas (Quadro 13).

Quadro 13 – Caracterização estatística dos dados utilizados no estudo.

	<i>MaxD</i> (mm)	<i>MinD</i> (mm)	<i>MDS</i> (mm)	<i>TGR</i> (mm)	<i>HSUM</i> (mm)	<i>RR</i> (mm)	<i>P</i> (mm)	<i>ET<sub>0</sub></i> (mm)	<i>K<sub>s</sub></i>
<b>Média</b>	79.859	79.624	0.235	0.007	203.816	0.950	--	3.762	0.778
<b>D.Padrão</b>	0.919	0.884	0.161	0.071	46.583	1.306	--	2.09	0.268
<b>Mínimo</b>	78.393	78.159	0.002	- 0.286	142	0.0	0.0	0.72	0.329
<b>Máximo</b>	81.609	81.577	0.890	0.538	313.8	5.271	21	7.57	1
<b>25%</b>	78.835	78.751	0.110	-0.02	150.475	--	--	1.795	--
<b>50%</b>	80.107	79.664	0.186	0.007	217.75	--	--	3.375	--
<b>75%</b>	80.419	80.118	0.326	0.036	241.75	--	--	5.655	--

O quadro permite obter algumas informações acerca das condições do ano. Uma média de precipitação de 0.669 mm/dia e uma média de *ET<sub>0</sub>* de 3.762 mm são concordantes com um ano e região onde as condições meteorológicas são de um clima bastante seco no Verão (Csa), sendo 50% dos valores de *ET<sub>0</sub>*, acima de 3.375 mm com um máximo de 7.57 mm. A precipitação é também bastante concentrada, por exemplo o valor máximo de 21 mm num único dia.

Construiu-se a matriz de relação entre as variáveis TGR, MDS, HSUM, *ET<sub>0</sub>* e *K<sub>s</sub>*, por representarem respectivamente, os índices derivados do dendrómetro (TGR e MDS), a dinâmica da água no solo pelo valor da sonda (HSUM), as condições meteorológicas (*ET<sub>0</sub>*) e de stress hídrico (*K<sub>s</sub>*) (Figura 24).

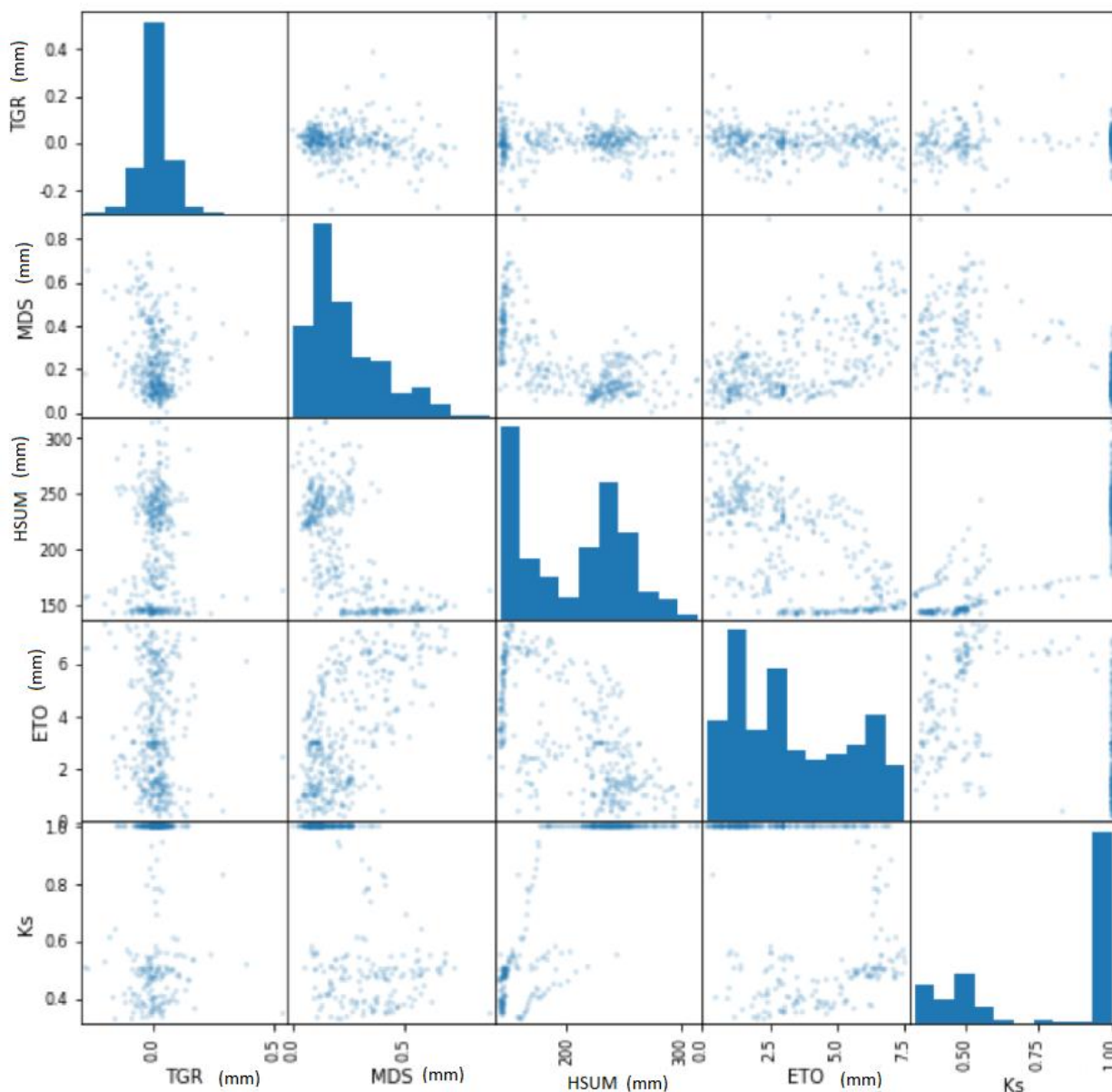


Figura 24 – Matriz de Relação das variáveis.

É possível através de uma análise visual dos dados, inferir que TGR não aparenta apresentar qualquer tipo de relação com as restantes variáveis, enquanto MDS aparenta variar com HSUM,  $ET_0$  e  $K_s$ , sugerindo que possa ser um índice interessante para detecção do stress.

Na figura 25 podemos ver a evolução de MDS ao longo dos dias do ano, e a distribuição dos seus valores num diagrama “boxplot” estando marcados a diferentes cores, os estados fenológicos, numerados de um a nove.

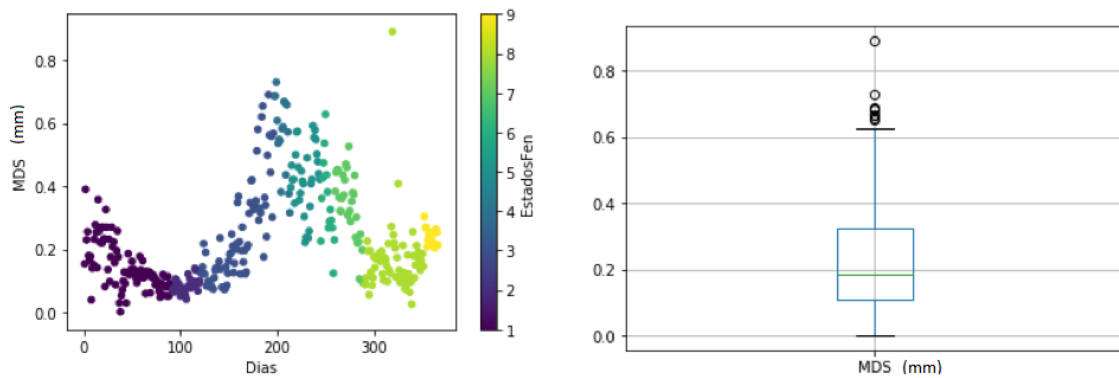


Figura 25 – Caracterização e Distribuição no tempo de MDS.

Os valores de MDS vão crescendo ao longo do ano, tendo um pico por volta do dia 200, e voltando a reduzir até ao final do ano com a evolução das fases fenológicas.

O diagrama “*boxplot*” mostra que existem alguns *outliers* mas a distribuição de valores concentra-se na região entre 0.1 e 0.3.

Na figura 26 a mesma abordagem foi usada para mostrar a evolução de TGR ao longo dos dias do ano.

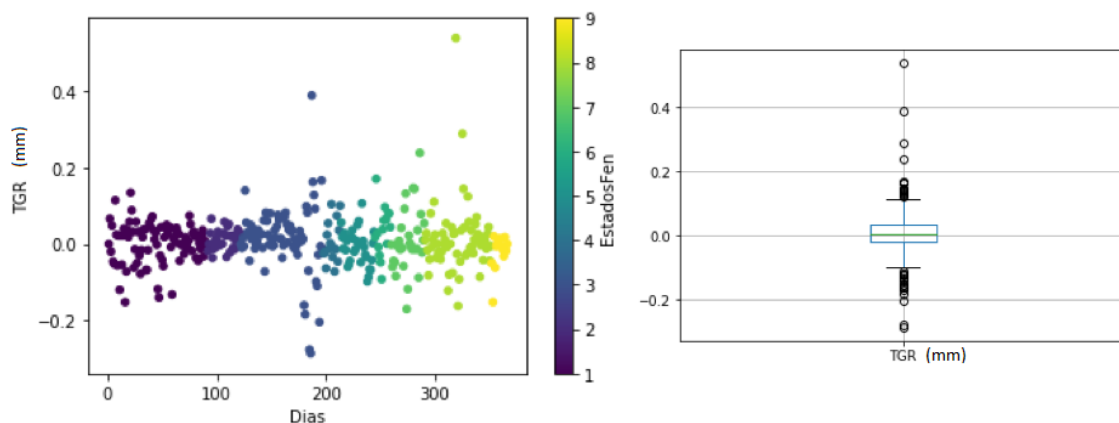


Figura 26 – Caracterização e Distribuição no tempo de TGR.

Verifica-se que TGR oscila constantemente ao longo do ano e em todas as fases do ciclo de crescimento, não existindo uma clara tendência de crescimento em qualquer parte do ano, mas sim uma grande alternância de dia para dia e com grandes amplitudes nos valores como se pode observar no diagrama “*boxplot*”.

Uma melhor interpretação da relação entre os dados obtidos pode ser visualizada na figura 27 onde é possível ver paralelamente, a evolução de HSUM, reserva do solo teórica,  $ET_a$ ,  $K_s$ , MDS e TGR ao longo do ano.

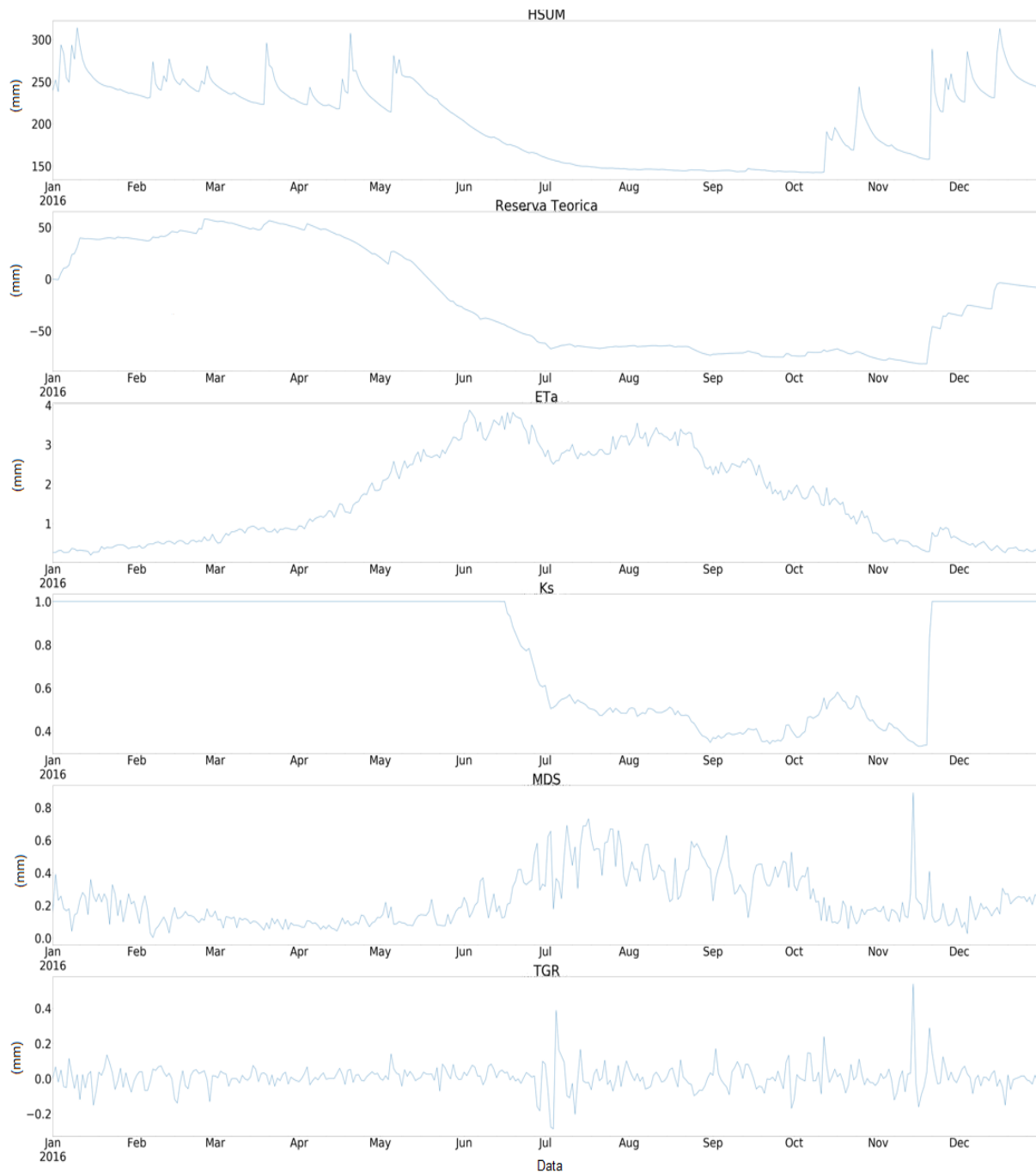


Figura 27 – Comparação dos diferentes elementos em estudo ao longo do tempo. De cima para baixo: ASUM, Reserva Teórica,  $ET_a$ ,  $K_s$ , MDS e TGR.

Neste gráfico é facilmente perceptível o aumento dos valores de MDS no início do período de stress, e a sua reacção a alterações no valor de  $K_s$ . Quanto a TGR, é impossível identificar uma relação ou resposta às variações do estado hídrico do solo. Realçar também que apesar de os valores absolutos de HSUM não serem iguais aos da reserva teórica, a sua evolução é muito semelhante, sendo a sonda mais sensível a variações.

Na figura 28 encontra-se ilustrado o período em que  $K_s < 1$ , onde é possível observar com maior detalhe a evolução dos valores medidos pela sonda, o índice MDS e  $K_s$  calculado nesse período.

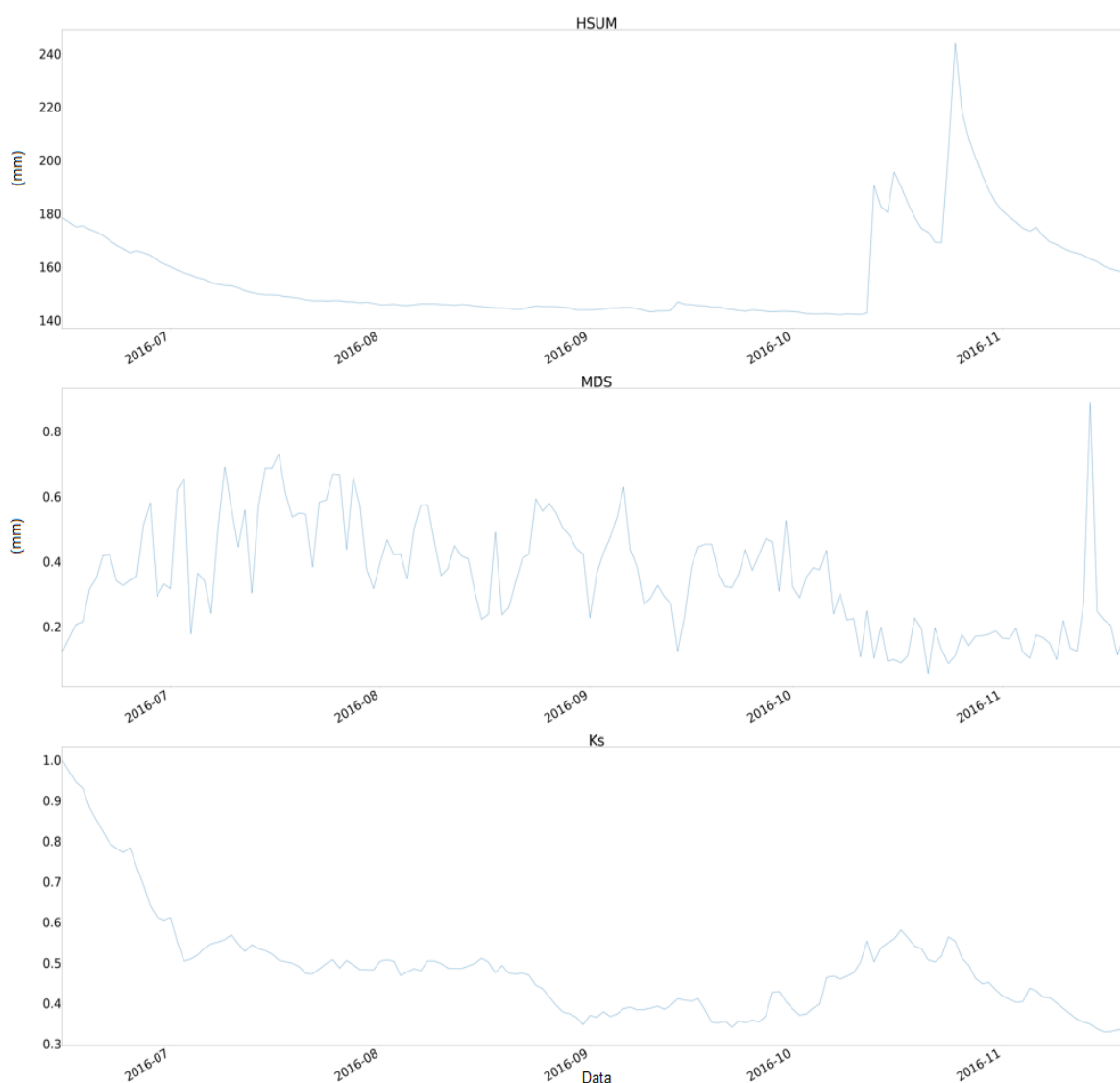


Figura 28 – Comparação de variáveis no período de stress estimado. De cima para baixo: HSUM, MDS e  $K_s$ .

A evolução da humidade do solo, parece estagnar ou decrescer a uma taxa muito ligeira durante este período, o que corresponde ao período de maior stress hídrico em que  $K_s$  é mais baixo. Também a amplitude e valor absoluto de MDS parece ser máximo durante este período. Com o aumento do teor de humidade do solo, nota-se uma diminuição no valor de MDS.

Para quantificar as diferenças entre os dois períodos analisados foi calculada a função densidade de probabilidades para a variável MDS nos dois períodos, bem como a média, desvio padrão, máximo e mínimo. Na figura 29 estão representadas as duas distribuições. O quadro 14 completa a informação gráfica.

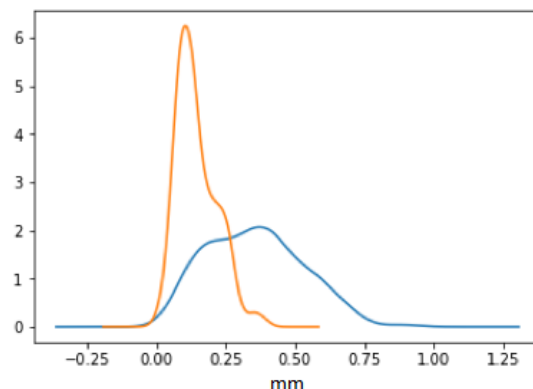


Figura 29 – Distribuição dos valores de MDS pelos períodos em estudo. A amarelo: Não Stress, a azul: Stress Estimado.

Quadro 14 – Distribuição dos valores de MDS (mm) nos dois períodos

	<b>Média</b>	<b>D.Padrão</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>25%</b>	<b>50%</b>	<b>75%</b>
<b>Sem Stress</b>	0.144	0.072	0.002	0.39	0.088	0.126	0.190
<b>Com Stress</b>	0.355	0.168	0.056	0.890	0.22	0.350	0.460

Denota-se que a distribuição dos valores para o período em que a planta não está em stress hídrico é mais baixa, sendo que 75% encontra-se abaixo de 0.19. Enquanto que para a zona de stress hídrico apenas 25% dos valores se encontram abaixo de 0.22, confirmando a existência valores comuns, mas sendo a maioria dos valores mais altos associados à zona de stress.

#### 4.1.2 – 2017

Para o ano de 2017, repetiu-se a mesma metodologia de análise aplicada aos dados de 2016, podendo deste modo reforçar ou refutar as conclusões da análise do ano anterior.

A caracterização geral dos dados sensoriais e da matriz de relação entre as variáveis encontram-se na figura 29 e no quadro 15.

Quadro 15 – Caracterização das variáveis recolhidas.

	MaxD(mm)	MinD(mm)	MDS(mm)	TGR(mm)	HSUM(mm)	RR(mm)	P(mm)	$ET_0$ (mm)	$K_s$
<b>Média</b>	88.226	87.983	0.243	0.0048	218.516	1.409	0.732	4.416	0.853
<b>D.Padrão</b>	0.540	0.526	0.169	0.092	48.165	1.117	3.342	1.738	0.234
<b>Mínimo</b>	86.453	85.943	0.036	-0.374	151.4	0.0	0.0	1.09	0.336
<b>Máximo</b>	88.831	88.340	1.222	0.804	343.5	4.127	34.2	7.79	1
<b>25%</b>	87.747	87.604	0.092	-0.022	169.92	--	--	2.922	--
<b>50%</b>	88.129	88.129	0.228	0.006	215.8	--	--	4.4	--
<b>75%</b>	88.340	88.340	0.335	0.028	257.3	--	--	5.807	--

Pelos dados apresentados, é possível constatar que o ano de 2017 apresentou uma média de valores de  $ET_0$  superior a 2016, tendo sido um ano com menos eventos de precipitação.

Na matriz de relações entre as variáveis, a relação entre os valores de MDS e a evolução dos dados da sonda, representados por HSUM, não é tão evidente como no ano de 2016. Por outro lado, observa-se a tendência para valores maiores de MDS com o aumento do  $ET_0$  como era expectável. Tal como em 2016 a variável TGR não apresenta qualquer relação tangível com as restantes variáveis do estudo (Figura 30).

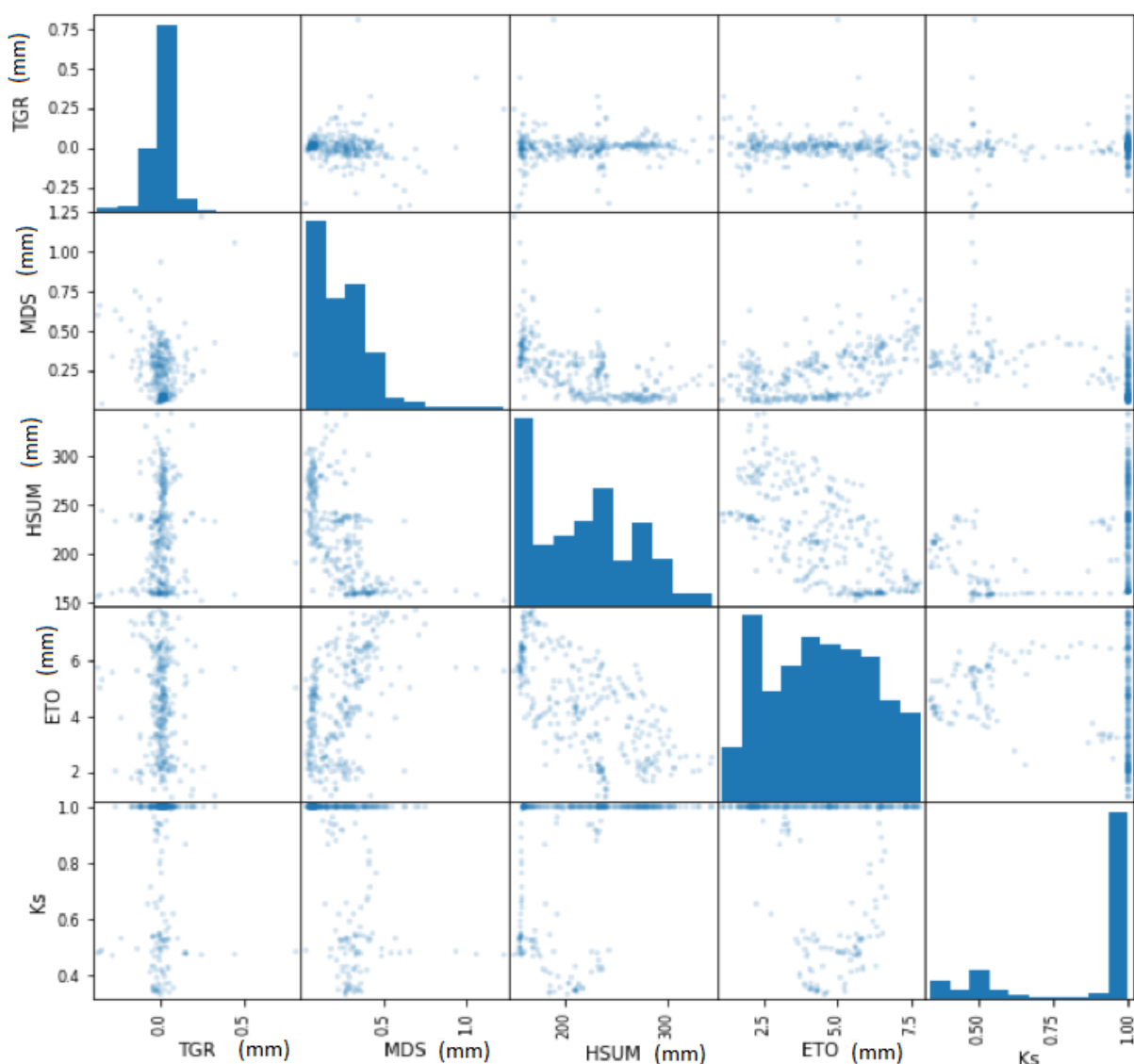


Figura 30 – Matriz de relação das variáveis em estudo.

Na figura 31 pode-se visualizar a evolução de MDS ao longo dos dias do ano, bem como a sua distribuição de valores no gráfico “*bloxplot*”.

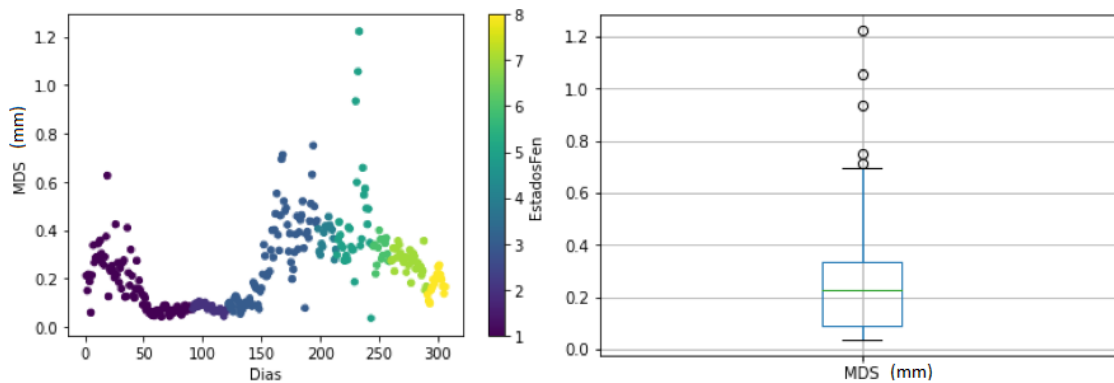


Figura 31 –Caracterização e distribuição de MDS no tempo.

Tal como em 2016 os valores de MDS evoluem ao longo do ano, subindo gradualmente, e voltando a descer no final do ano, existindo neste ano um maior número de *outliers*, visível no diagrama de “*boxplot*”. A distribuição mostra que os valores são ligeiramente mais elevados que no ano anterior.

A evolução de TGR ao longo do ano e a distribuição dos seus valores encontram-se na figura 32.

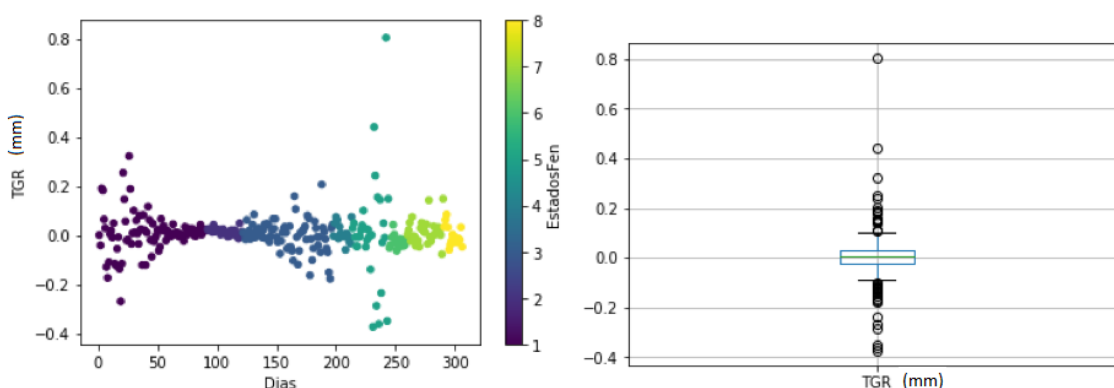


Figura 32 – Caracterização e Distribuição de TGR no tempo.

Mais uma vez os valores de TGR sofrem pequenas oscilações, sem apresentarem um período em que se vislumbre claramente uma tendência de valores positivos ou negativos. Os *outliers* são em maior quantidade do que no caso de MDS.

Seguindo o mesmo procedimento para a análise, a figura 33 apresenta a evolução, em paralelo de HSUM, Reserva Teórica,  $ET_a$ ,  $K_s$ , MDS e TGR.



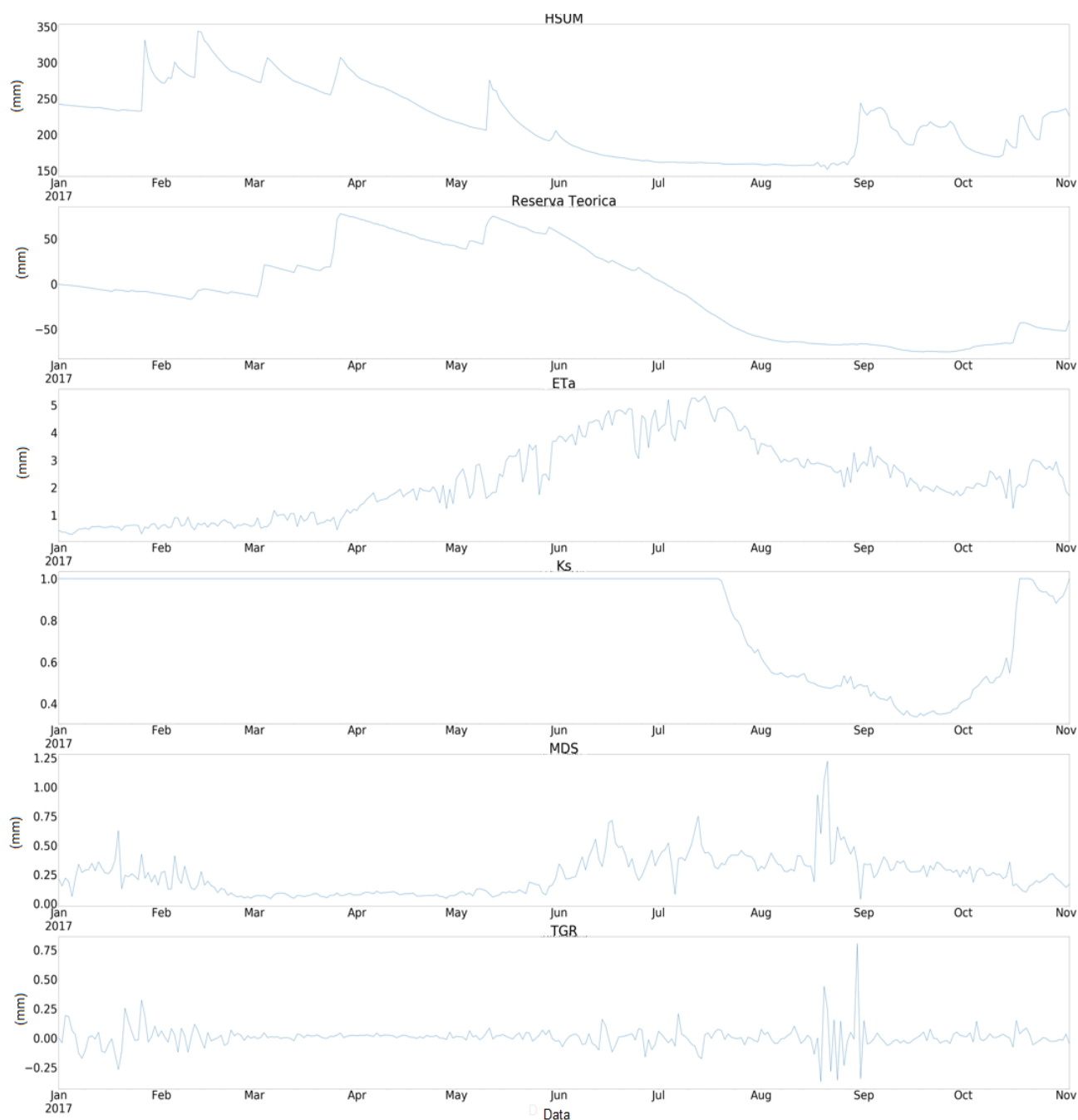


Figura 33 – Comparação das variáveis em estudo ao longo do tempo. De cima para baixo: HSUM, Reserva Teórica,  $ET_a$ ,  $K_s$ , MDS e TGR.

Neste gráfico a relação entre a evolução do stress hídrico ( $K_s$ ) e os valores de MDS aparenta estar menos correlacionada do que no ano anterior, e mais uma vez a evolução da reserva teórica segue a evolução dos valores de HSUM, sendo estes mais sensíveis às oscilações do conteúdo de água do solo, como é expectável. O aumento nos valores de MDS acontece antes da entrada em período de stress aparente sinalizada pelo valor de  $K_s$ .

Para este ano apresentam-se os gráficos do valor de MDS no período de stress sinalizado por  $K_s$  e no restante período do ano (Figura 34).

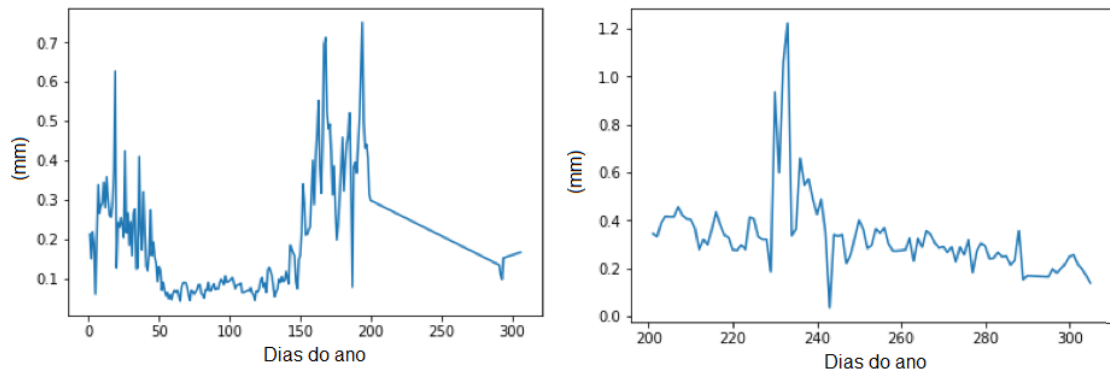


Figura 34 – Valores no período de Não Stress / Stress. À direita - Período de Stress, À esquerda- Período de Não Stress

Claramente denota-se que o aumento nos valores de MDS acontece antes do período de stress previsto pelo balanço hídrico.

Na figura 35 apresenta-se a evolução dos valores de HSUM, do MDS, e de  $K_s$ .

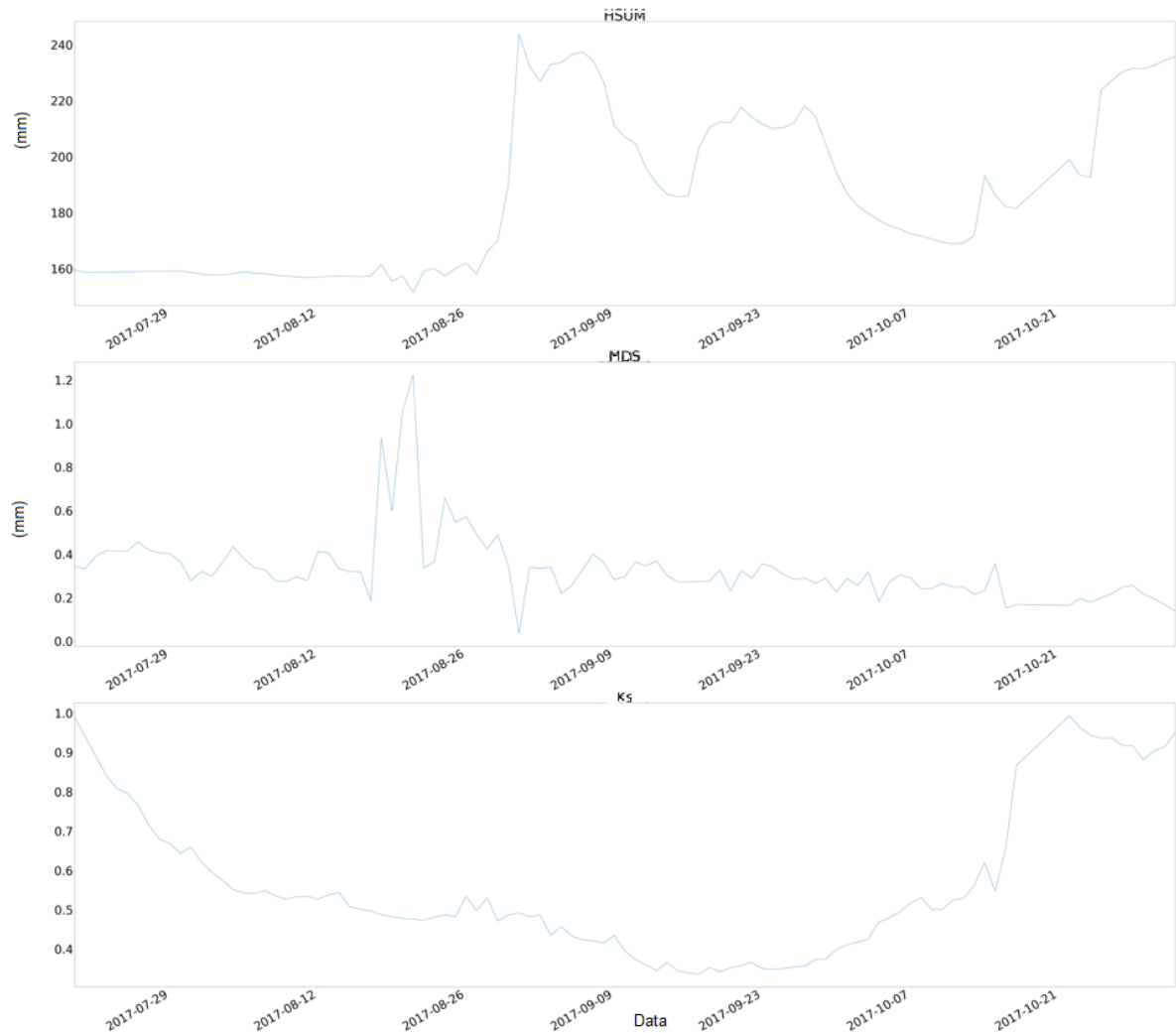


Figura 35 – Valores de HSUM, MDS e  $K_s$  no período estimado de stress.

Ao contrário do que acontece no ano de 2016, em 2017 mudanças bruscas na amplitude dos valores não se reflectem nos valores de MDS apesar de estes se manterem constantemente mais elevados do que noutras alturas do ano.

A figura 36 representa a densidade de probabilidade para os dois períodos do ano em comparação.

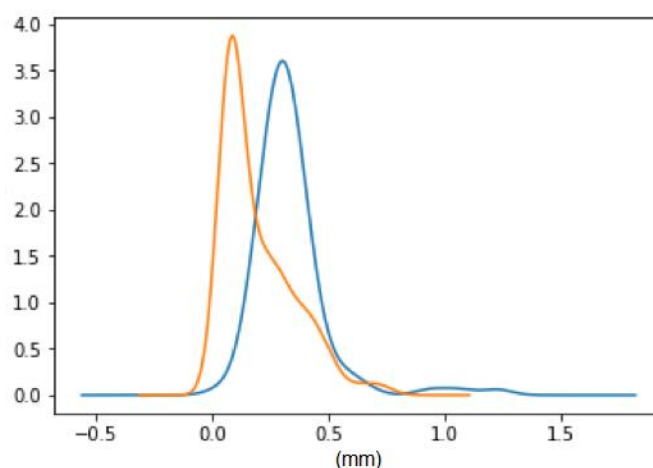


Figura 36 – Distribuição dos valores de MDS no período de não stress e de stress. A amarelo – Não Stress. A Azul- Stress.

Neste caso e como se previa, a distribuição dos valores é muito mais próxima do que no ano de 2016, havendo uma maior sobreposição das duas distribuições.

O quadro 16 apresenta estes valores em maior detalhe.

Quadro 16 – Distribuição de MDS (mm) para os períodos em estudo.

	Média	D.Padrão	Mínimo	Máximo	25%	50%	75%
<b>Sem Stress</b>	0.196	0.152	0.04	0.75	0.076	0.128	0.288
<b>Com Stress</b>	0.335	0.163	0.036	1.22	0.256	0.306	0.364

Pela tabela é possível novamente verificar uma diferença significativa entre os valores médios de cada zona em estudo. Para além disso, será de realçar que para a zona do ano sem stress aparente, 75% dos valores de MDS estão abaixo de 0.288, o que, apesar de mais elevado que no ano anterior, continua a ser representativo do efeito do stress hídrico na evolução dos valores do índice.

Comparando os valores dos dois anos (quadro 17), é possível pressupor um valor limite para considerar como indicador de stress hídrico a partir das medidas do dendrómetro.

Quadro 17 – Selecção de um valor base de detecção de Stress Hídrico a partir de MDS(mm).

Média	Com Stress	Sem Stress
<b>2016</b>	0.355	0.144
<b>2017</b>	0.335	0.196
<b>Indicador</b>	0.4	< 0.4

Devido à oscilação dos valores apresentada em 2017, faz sentido dar alguma margem ao valor selecionado como indicador, sendo fixado em 0.4. Mesmo assim é possível constatar que aplicando esta metodologia aos dados de 2017, o período de stress seria maior do que o considerado teoricamente (Figura 37).

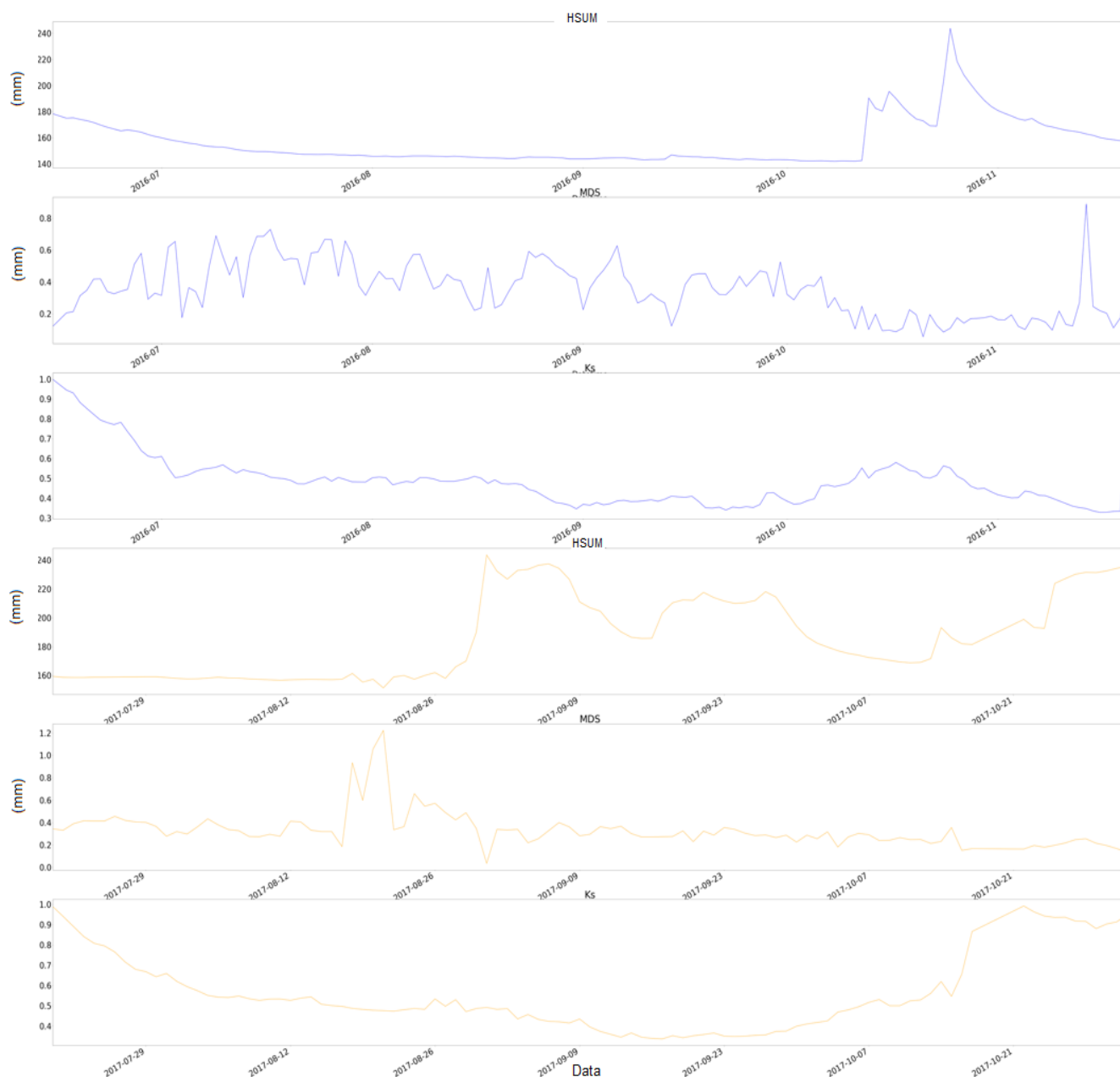


Figura 37– Comparação dos anos 2016 e 2017 para as variáveis HSUM, MDS e  $K_s$ , nos períodos de stress hídrico estimado. 2016 a azul e 2017 a amarelo

Na figura observa-se que o período considerado para a planta se encontrar em stress é mais pequeno do que se considerarmos apenas o indicado pelo  $K_s$ , mas que este começa e

termina mais cedo, coincidindo melhor com a zona em que a reserva do solo está com maior depleção, de acordo com os dados da sonda.

## 4.2 – Algoritmos ML de previsão do valor de MDS

Como MDS é o índice mais relacionado com as alterações nos conteúdos de água do solo e os períodos de stress descritos no capítulo anterior, foi selecionado como o indicador mais relevante a obter a partir de um conjunto de dados simples, pois este tipo de previsão pode substituir alguns dos sensores presentes no campo, tornando a informação mais acessível ao pequeno agricultor, sem a necessidade de adquirir sensores de dendrometria.

O algoritmo de floresta de decisão aleatória (FDA) implementado neste trabalho utiliza as variáveis descritas no quadro 18 como preditores de MDS.

Quadro 18 – Variáveis de previsão e sua amplitude de valores.

Variável	Unidades
$ET_0$	mm
$\Delta S$	mm
HSUM (Reserva de água no solo)	[0-1], normalizada
RR (Registos de Rega)	mm
P (Precipitação)	mm
Ano	[2015-2017]
Mês	[1-12]
Dia	[1-31]

Foram utilizadas 5000 árvores de decisão individuais para construir o algoritmo. Os dados foram separados em 75% para treino do modelo e 25% para validação do mesmo.

Após o período de treino calculou-se a margem de erro média das previsões nos dados de teste e a percentagem de valores previstos correctamente, estando ambas as métricas no quadro 19.

Quadro 19 – Métricas de Avaliação da FDA.

Métrica	Valor
Desvio Médio por previsão	0.0548
% Valores correctos	70.89

Os valores apresentados indicam que o algoritmo se comporta razoavelmente bem na tarefa de prever os valores de MDS, uma vez que um desvio médio por previsão de 0.0548 numa variável que varia na casa das algumas décimas é aceitável, e a previsão correcta de 70% dos valores do conjunto de validação é também uma boa estimativa tratando-se de um problema de

regressão numa variável contínua. Na figura 38 pode ser visualizado um diagrama geral de uma das árvores de decisão da FDA.



Figura 38 – Vista Geral de uma das árvores do algoritmo FDA.

Um factor importante a avaliar é a influência de cada variável de previsão no resultado final, podendo a partir desta análise reduzir o número de variáveis utilizadas. Na figura 39 estes resultados estão representados em gráfico de barras.

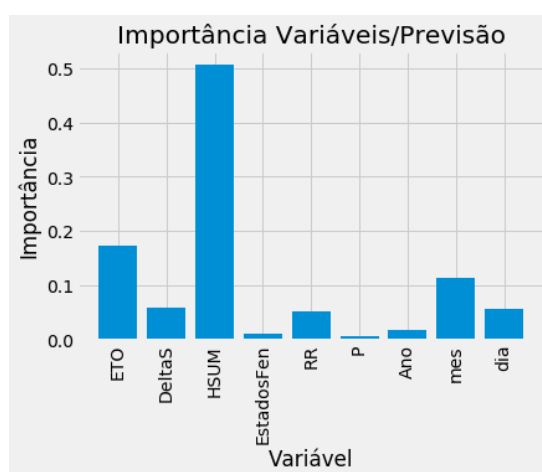


Figura 39 – Importância das variáveis no resultado final do treino.

Como se pode observar, os valores de HSUM são os que mais contribuem para a selecção do valor de MDS, confirmando a relação entre as duas variáveis.  $ET_0$  aparece como segundo factor mais importante, seguido do mês do ano e da variação diária da reserva. Todas as variáveis apresentam algum grau de importância e logo contribuem para a qualidade do modelo não podendo ser retiradas.

Na figura 40 podemos observar graficamente a qualidade do modelo desenhando os valores reais e os valores previstos, ordenados pela data como ponto de referência.

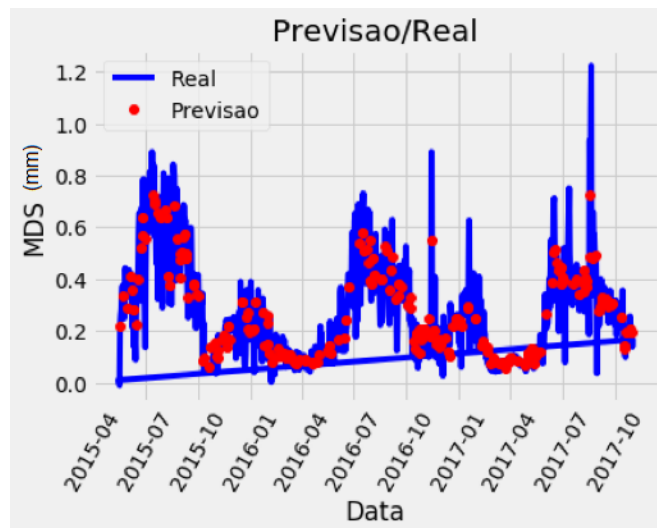


Figura 40 – Dados Reais e Pontos previstos no conjunto de teste

A correlação entre os valores reais e os valores previstos é bem visível no gráfico, enfatizando os valores obtidos no quadro 18 sobre o erro das estimativas.

O modelo de rede neuronal (ANN), implementado e descrito no capítulo 3, para estimar os valores de MDS foi implementado com 2 camadas escondidas, a primeira de 18 perceptrões e a segunda de 9.

O processo de treino foi conduzido em 260 “*Epochs*”, sendo que cada “*Epoch*” corresponde a uma passagem de todos os dados de treino pela rede neuronal. Como o número total de dados é reduzido, é natural que o número de *epochs* até a convergência seja baixo. A partir desse ponto o algoritmo já não melhora a sua performance como é observável na figura 41.

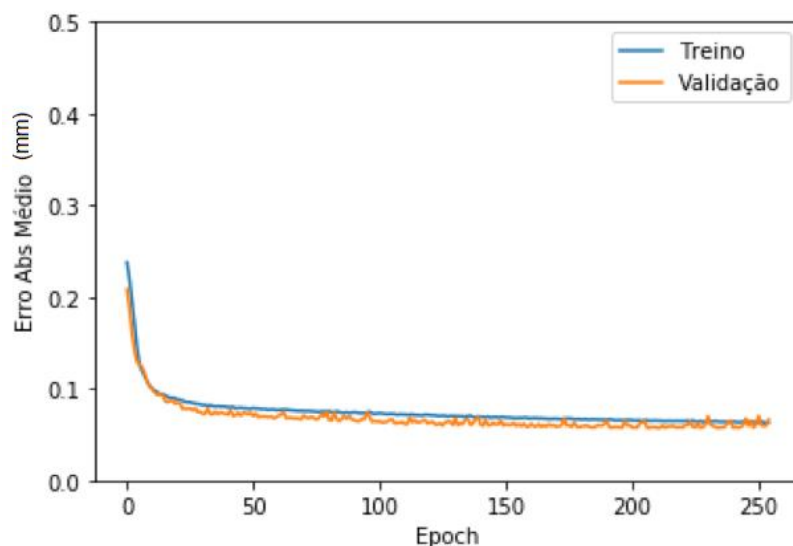


Figura 41 – Evolução do erro absoluto médio ao longo do processo de treino da ANN

A média dos erros obtidos para o conjunto de dados de teste é de 0.0624 cm, um pouco acima do obtido para o mesmo conjunto de dados com a FDA. A precisão das previsões para o conjunto de teste é de 67.37 %, também um pouco abaixo do obtido para a FDA. Na figura 42 é possível observar os valores previstos pelo algoritmo e os valores reais.

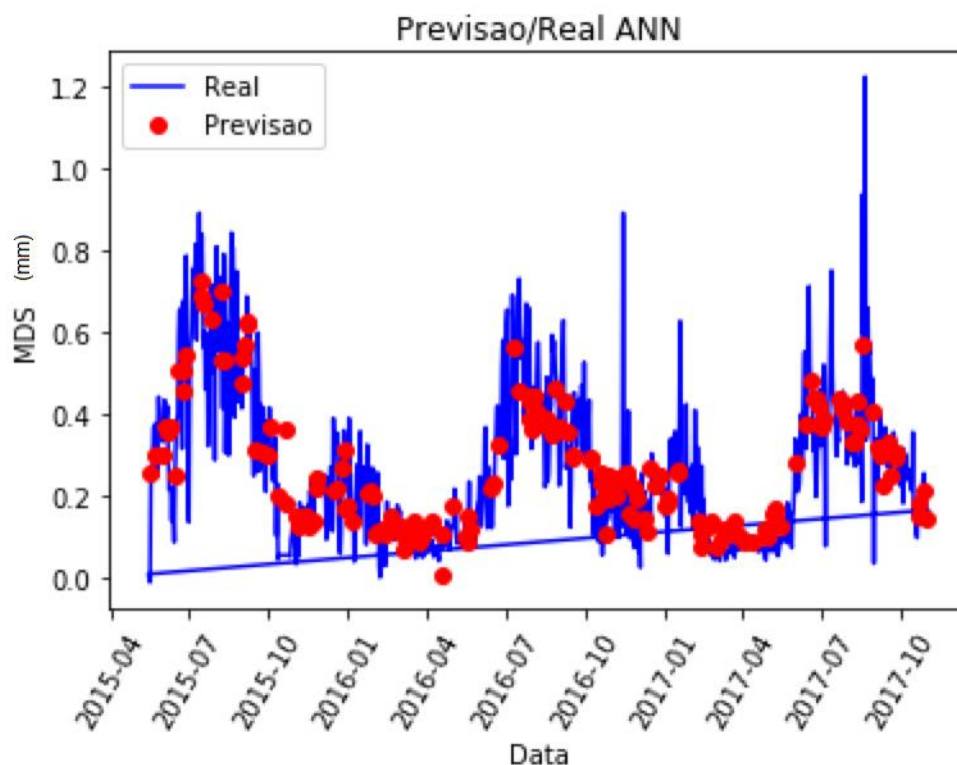


Figura 42 – Dados reais e pontos previstos pela ANN no conjunto de dados de teste.

Graficamente é observável que o algoritmo prevê a maior parte dos valores correctamente, sendo que é notória a diferença entre os dois algoritmos para o caso em que os valores são de uma grandeza muito díspar em relação à média. Nesse caso o algoritmo FDA obtém melhores resultados que o ANN.

Uma comparação entre os erros observados na classificação do conjunto de teste fornece uma visão mais detalhada da qualidade dos dois algoritmos. A figura 43 apresenta estes dados em formato de histograma.



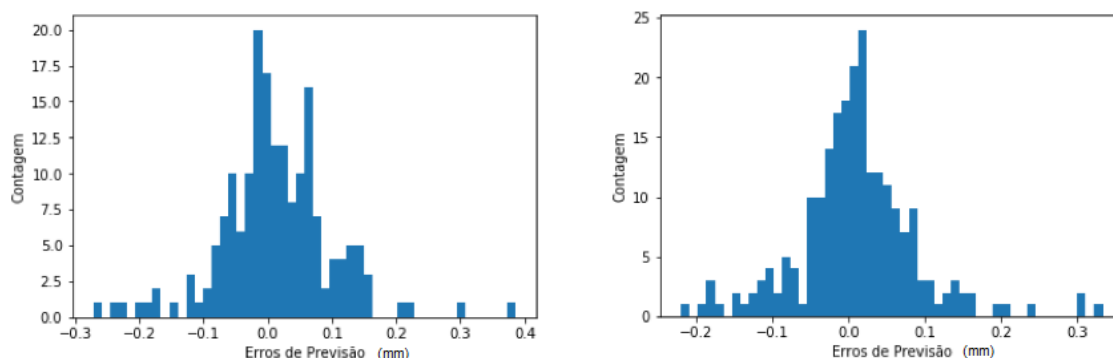


Figura 43 – Histogramas dos erros de previsão de ambos os algoritmos. Esquerda – ANN, Direita – FDA.

O número de previsões com erro próximo de 0 é maior para FDA do que para ANN. No sentido contrário os erros de maior amplitude pertencem às previsões de ANN (entre -0.3 mm e 0.4 mm, vs -0.2 mm a 0.3 mm).

### 4.3 – Gestão inteligente de rega utilizando AR

O algoritmo DQN detalhado no capítulo 3 foi implementado com as parametrizações detalhadas no quadro 20.

Quadro 20 – Parametrizações do sistema de aprendizagem reforçada

Parâmetro	Valor
Número de episódios de treino	1000
Número de iterações/episódio	365
Valor de desconto de recompensas futuras	0.85
Taxa de aprendizagem	0.005
Tamanho da memória(Iterações)	2000

As duas redes neurais utilizadas apresentam a mesma estrutura, sendo compostas por duas camadas de 24 perceptrões cada, e uma camada de saída de 10 perceptrões, correspondendo cada um ao valor Q do par (estado,acção) para cada uma das acções possíveis.

O número elevado de episódios conduziu a um processo de treino de cerca de 10 horas num computador portátil normal. No entanto, este número garante que o algoritmo simulou todas as estratégias possíveis de maximização da recompensa final e tenha completado o maior número de episódios possível. A duração de um episódio foi escolhida de acordo com os dias de um ano, garantindo que simulamos um ciclo completo da cultura. A cada episódio gravou-se a recompensa total, as acções tomadas em cada episódio e a evolução da reserva simulada. Um resumo dos valores acumulados das recompensas é apresentado na figura 44.

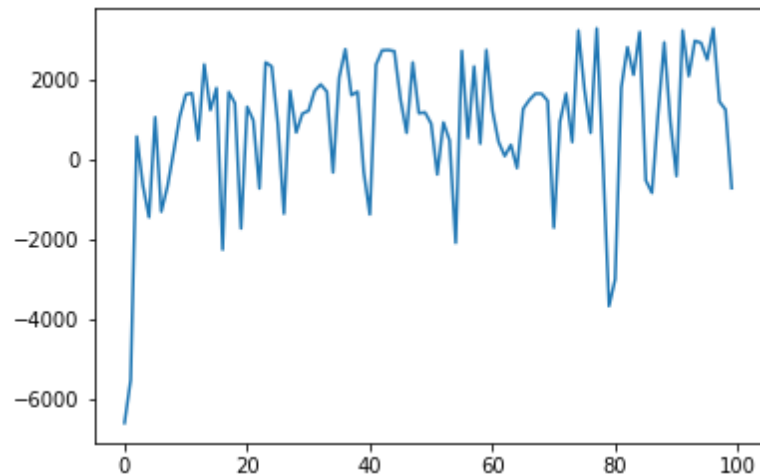


Figura 44 – Recompensa de cada episódio ao longo do processo de treino. Os valores apresentados são a recompensa total, obtida a partir do somatório das recompensas individuais obtidas pela passagem em cada estado durante um episódio.

Como é possível constatar, a recompensa sofre oscilações bruscas ao longo do processo de treino, sendo que é possível verificar uma tendência crescente no valor da recompensa à medida que o algoritmo vai aprendendo a escolher as melhores acções em cada estado encontrado.

Para avaliar a prestação do algoritmo tradicionalmente calcula-se a média ao longo do tempo da recompensa acumulada. Neste caso os objectivos da simulação são a conclusão do ciclo sem entrar nos estados F (representam situação de stress extremo) e a distribuição dos eventos de rega ao longo do ano para podermos comparar qual a diferença entre as sugestões do algoritmo e a rega efectuada pelo agricultor.

Tendo em consideração apenas as simulações terminadas com sucesso (quase a totalidade) e representando a que atingiu o maior valor de recompensa podemos avaliar a estratégia de rega seguida pelo algoritmo e a do agricultor. Na figura 45 apresenta-se essa comparação ao longo dos dias do ano.

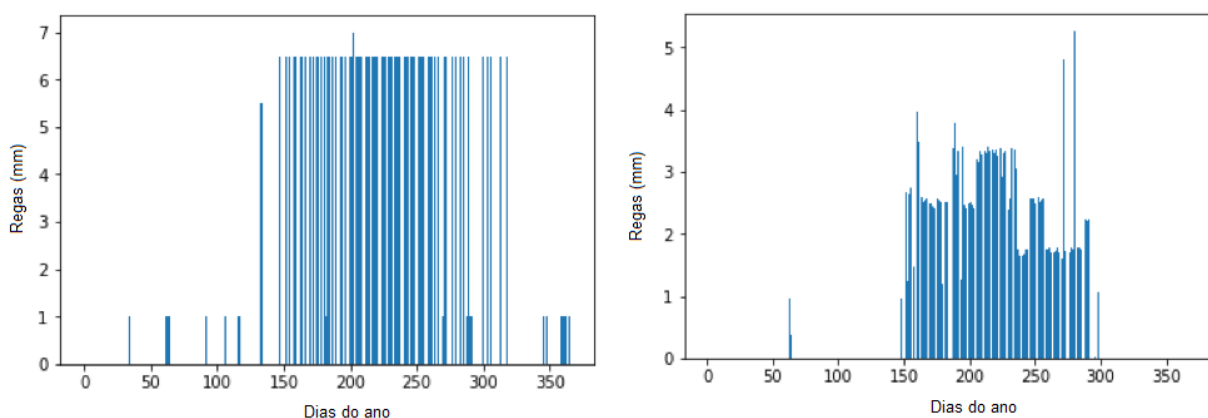


Figura 45 – Eventos de Rega Simulados e os que efectivamente foram realizados pelo agricultor. Direita- Agricultor, Esquerda - Simulação

Ao analisar a figura verifica-se que o período de regas é maior no caso do algoritmo, regando mais dias do ano e com maior amplitude em cada evento. No entanto o período de regas mais frequentes situa-se entre os dias 150 e 300 em ambos os casos, como era expectável, sendo este o período de verão.

Para além destes dados, a figura 46 apresenta a evolução da reserva simulada, sendo este o principal parâmetro avaliado nas directivas do algoritmo e cujos valores representam o sucesso ou insucesso da simulação.

Verificamos que os valores da reserva se mantêm na zona de conforto durante o início e o fim do ano, sendo de especial atenção o facto de as regas seleccionadas pelo algoritmo manterem a reserva quase sempre acima de 60% mesmo no pico do verão, reflectindo as boas escolhas do algoritmo, quer na intensidade, quer nos intervalos de tempo escolhidos para a rega.

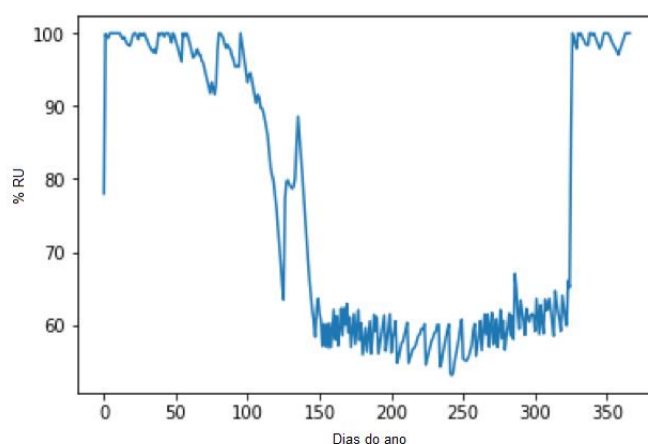


Figura 46 – Variação da reserva de água do solo simulada para o episódio de treino de maior recompensa.

## 5 – DISCUSSÃO

A análise aos dados de dendrometria recolhidos durante os anos de 2016 e 2017 permitiu concluir que o índice MDS tem potencial para ser utilizado num olival do tipo superintensivo como um bom indicador do estado hídrico das plantas. Partimos do pressuposto de que, como já foi referido várias vezes neste trabalho, uma verdadeira gestão inteligente e eficiente da rega necessita de indicadores da planta robustos para complementar o maior número de informações que possamos retirar do solo.

No ensaio realizado, verificaram-se algumas oscilações nos valores medidos, especialmente para o ano de 2017. Este tipo de oscilação é expectável, pois, ao longo de 2017, o sensor foi ajustado durante a campanha, pelo menos duas vezes, introduzindo ruído e incerteza nas medições. Para além disso, a evolução da reserva de água disponível no solo foi mais desfavorável do que em 2016. Nos dados de 2016, consegue-se identificar também um período rápido de crescimento do tronco entre Março e Junho, correspondendo ao período de crescimento vegetativo, e uma estagnação durante os meses de Verão. Tal relação indica que a planta em estudo se encontra saudável e segue o seu ciclo normal de desenvolvimento.

Uma das principais ilações a retirar deste trabalho é que, em ambos os anos, foi possível utilizar uma estimativa de  $K_s$  para conseguir facilmente separar os dois conjuntos de valores de MDS, apresentando estes algumas características diferentes em amplitude e valor absoluto. Tal não era esperado, uma vez que a maioria dos estudos feitos em olival até agora, indicam que, para obter alguma relação entre os valores de MDS e os efeitos da disponibilidade hídrica da planta, seria necessário utilizar uma planta de referência em conforto hídrico. Para além disso, no ano de 2017, a discrepância entre os períodos de stress hídrico seleccionados pelo indicador MDS ou pelo valor de  $K_s$  empírico, sugerem que o valor de MDS possa ter utilidade na identificação antecipada da entrada da planta num período de stress. Os próprios dados da sonda de humidade do solo são consistentes com um período de baixo teor de água no solo. Estas observações permitem inferir que o período de stress identificado a partir do valor de  $K_s$  empírico seja inferior àquele a que efectivamente a planta esteve sujeita, estando este período de stress melhor definido pela mudança de amplitude nos valores de MDS.

Esta análise dos dados de dois ciclos culturais procura abrir a possibilidade de num novo estudo, com indicadores mais precisos do estado hídrico da planta, como o potencial hídrico foliar, validar o potencial de utilização de MDS como indicador do estado hídrico, por si só, e não estando dependente de valores de referência. No entanto é importante reforçar que a abordagem aqui seguida é apenas indicativa para este tipo de olival superintensivo e não deve ser extrapolada para outras culturas ou condições de cultivo. Para além disso, a definição dos valores de  $K_s$  seguiu a abordagem empírica descrita na FAO 56 e apresenta algumas limitações, não devendo ser considerados 100% fidedignos os períodos de stress identificados por esta via e que serviram de base a esta parte do trabalho.

Como complemento a este trabalho de análise, foi também criado um algoritmo que permite prever o valor de MDS com base em inputs sensoriais de fácil aquisição no campo, uma vez demonstrado o seu potencial como indicador de stress. Foram utilizadas duas abordagens para a criação do algoritmo, uma de ML clássica, com a utilização de uma Floresta de Decisão Aleatória (FDA) e uma Rede Neuronal.

Os resultados obtidos pela FDA apresentam maior precisão na previsão dos valores de MDS ao longo do conjunto de dados em estudo do que os obtidos pela rede neuronal (ANN). Tal facto pode estar relacionado com o baixo volume de dados para o treino satisfatório de uma rede neuronal ou com a própria estrutura escolhida para essa mesma rede. De realçar que no período do ensaio foram testadas várias configurações da rede, optando-se pela que apresentou melhores resultados de classificação. Para dados escassos e problemáticas onde é difícil definir uma estrutura “à priori”, a FDA pode ser tida como uma excelente opção para aplicações de ML com dados sensoriais do campo.

Estes resultados demonstram que é possível construir algoritmos que cumprem os objectivos sem demasiada complexidade e com um número limitado de variáveis. A margem de erro nas previsões do conjunto de dados de teste é encorajadora tendo em conta que diariamente o valor de MDS pode variar algumas décimas e o erro se encontra em cerca de meia décima.

Considerando o conjunto de análises do subcapítulo 4.1, é mais uma vez evidente a relação entre MDS e a evolução da reserva de água no solo, representando esta variável mais de 50% da importância na selecção de valores obtidos na FDA. O valor de  $ET_0$  é também um factor de elevada relevância pois traduz as condições meteorológicas. Numa primeira abordagem não foi tida em consideração a introdução dos dias do ano como variável. No entanto a sua introdução aumentou a precisão das previsões. Uma explicação possível é que estas variáveis ajudam a traduzir alguns dos factores que afectam MDS e que não são possíveis de prever a partir dos inputs do solo e do clima, como os ciclos de crescimento, muito dependentes do calendário anual, bem como ciclos de fertilização e presença de pragas.

A capacidade de prever indicadores de stress derivados de variáveis simples representa uma oportunidade para a disponibilização de informação ao produtor a baixo custo e com um moderado grau de precisão. O tipo de abordagem utilizado neste estudo com o MDS pode ser aplicado a outras variáveis de relevância para a condução das culturas e permite a construção de um banco de dados com relevância para outras aplicações.

No entanto, o resultado não pode ser generalizado pois as condições edafoclimáticas e culturais são bastante específicas e estes dados não incluem a variabilidade necessária para uma adopção do algoritmo a nível global.

O algoritmo de FDA permitiu a construção de um sistema de aprendizagem reforçada onde este mesmo algoritmo permite simular as respostas da planta à aplicação de diversos volumes de rega.

O sistema utiliza a avaliação dos valores de MDS, variáveis meteorológicas e simulações da evolução da reserva hídrica do solo para fornecer recomendações da quantidade de água a aplicar por cada dia do ciclo. Um algoritmo desta complexidade envolve um enorme número de variáveis, desde logo um domínio sobre as técnicas de redes neuronais e aprendizagem reforçada que é muito difícil de adquirir sem consultar especialistas.

O algoritmo de RL obteve resultados razoáveis para a experiência em que foi inserido. O principal objectivo era conduzir um ciclo anual da cultura sem esta entrar em stress, e esse objectivo, após alguns ajustes nas recompensas e no ambiente de testes, foi conseguido. Durante este processo é de realçar a importância na escolha das recompensas para cada acção. Numa primeira iteração, o algoritmo era recompensado sempre que completava os 365 dias da simulação. Isto levou a que a sua estratégia fosse regar todos os dias do ano, mantendo a reserva a 100% o ano todo. Alterar a distribuição das recompensas permitiu evitar este comportamento, conduzindo a um algoritmo estável, que reage às alterações no ambiente de forma correcta. Este é um passo importantíssimo para a investigação de novas técnicas de gestão da rega e avaliação de parâmetros. Ao simular correctamente um ambiente, podemos cruzar informação de diferentes fontes e ajustar a estratégia de rega, simulando-a primeiro antes de a aplicar em vez de esperar pelos seus resultados na prática. Para além disso, o facto de nesta simulação se utilizar o algoritmo de FDA como resposta da planta realça ainda mais as potencialidades destas metodologias.

Um estudo sobre estes sistemas de aprendizagem permite averiguar e testar as reacções da planta a diferentes estratégias de gestão de rega com alguma confiança nos resultados obtidos. A vantagem de simular este tipo de cenários, como já foi referido anteriormente é a quantidade de informação que se pode obter para depois voltar a testar no campo em condições reais. Partindo de um resultado simulado é muito mais fácil eliminar cenários que não necessitam de ser testados em ambiente real, uma vez que ensaios de campo são dispendiosos e muitas vezes inconclusivos.

Neste trabalho, um dos objectivos era demonstrar que tais simulações são possíveis, estando, no entanto, o seu resultado condicionado pela falta de experiência nas técnicas em questão.

Com uma boa base de treino, mais dados e ajustes ao modelo podemos, no limite, obter um sistema que controle a rega de forma autónoma, adaptando a aplicação de água às diferentes variáveis que vai adquirindo na sua rede sensorial. Para além disso a abordagem pode ser utilizada noutras investigações como a calibração dos valores de  $K_c$  às condições locais das culturas ou à correcção dos valores de  $p$ , identificando períodos de stress hídrico.

## 6 – CONCLUSÕES

Os resultados obtidos neste trabalho serviram como base para o desenvolvimento de novas metodologias de análise e investigação de dados recolhidos no campo, criando alternativas para a valorização desses dados e automatização de soluções para a gestão da rega. Para o demonstrar usou-se a análise de dados de dendrometria o que permitiu validar a utilização do índice de variação de diâmetro do tronco (MDS) como tendo potencial para ser um bom indicador do estado hídrico das plantas neste caso em particular. Este estudo abre a possibilidade de num novo ensaio, com indicadores mais precisos do estado hídrico da planta, confirmar estes resultados.

Deve também ser dada ênfase às possibilidades que a análise de dados, de Big Data, fornece para a elaboração de novas metodologias de análise a partir das mais variadas fontes de dados no meio agrícola.

Para dados escassos e problemáticas onde é difícil definir uma estrutura “à priori”, os algoritmos do tipo floresta de decisão aleatória (FDA) devem ser tidos como uma excelente opção para aplicações de *machine learning* (ML) com dados sensoriais do campo.

Qualquer que seja o método, ou técnica escolhida para obter um algoritmo de previsão em agronomia, é importante ter sempre em consideração que qualquer resultado será apenas relevante globalmente se estivermos perante uma enorme quantidade de dados que sejam representativos das mais variadas condições meteorológicas, tipos de solos e variedades dentro da mesma cultura.

Neste trabalho, um dos outros objectivos era demonstrar que simulações com algoritmos de aprendizagem reforçada são exequíveis. Abordando este sistema como uma prova de conceito, verificamos que é possível a sua construção e obtenção de resultados não otimizados sendo claramente necessários ajustes aos algoritmos experimentais. Ao mesmo tempo é

possível afirmar que o potencial oferecido por estes sistemas é tremendo, uma vez que estes permitem reagir a alterações ambientais e culturais em tempo real, ajustando a sua resposta.

Futuramente, este trabalho deve ser complementado com vários ensaios em campo experimental, em locais onde exista um elevado histórico de dados de produção da mesma cultura. Este trabalho deve ser executado por uma equipa composta por elementos com experiência em ML e agrónomos que possam identificar as variáveis mais relevantes às problemáticas a estudar. Apenas este trabalho conjunto pode fornecer a robustez e qualidade necessária para que estes algoritmos sejam verdadeiramente aceites como uma ferramenta de trabalho ao dispor, quer de investigadores, quer de produtores que cada vez mais necessitam de adquirir conhecimento sobre as melhores práticas agrícolas, num mundo em constante mudança, onde a sua margem de erro é cada vez menor.

## 7 – REFERÊNCIAS BIBLIOGRÁFICAS

- Abraham, A. (2015) *Artificial Neural Networks*. Wiley.
- Adnan, M., Nazir, M. and Latif, M. A. (2017) 'Estimating Evapotranspiration using Machine Learning Techniques', 8(9), pp. 108–113.
- Allen, R. G. *et al.* (1998) 'Crop evapotranspiration - Guidelines for computing crop water requirements - FAO 56', *FAO - Food and Agriculture Organization of the United Nations*.
- Allen, R. G. *et al.* (2005) 'FAO-56 Dual Crop Coefficient Method for Estimating Evaporation from Soil and Application Extensions', *Journal of Irrigation and Drainage Engineering*, 131(1), pp. 2–13. doi: 10.1061/(ASCE)0733-9437(2005)131:1(2).
- Allen, R. G. and Pereira, L. S. (2009) 'Estimating crop coefficients from fraction of ground cover and height', *Irrigation Science*, 28(1), pp. 17–34. doi: 10.1007/s00271-009-0182-z.
- Bittelli, M. (2011) 'Measuring soil water content: A review', *HortTechnology*, 21(3), pp. 293–300.
- Campos, I. *et al.* (2010) 'Assessing satellite-based basal crop coefficients for irrigated grapes (*Vitis vinifera* L.)', *Agricultural Water Management*. Elsevier B.V., 98(1), pp. 45–54. doi: 10.1016/j.agwat.2010.07.011.
- Capraro, F., Patifio, D. and Tosetti, S. (2008) 'Neural Network-Based Irrigation Control for Precision Agriculture', *IEEE International Conference on Networking, Sensing and Control*, pp. 357–362.
- Carmo, M. *et al.* (2017) 'The N-P-K soil nutrient balance of Portuguese cropland in the 1950s: The transition from organic to chemical fertilization', *Scientific Reports*, 7(1), pp. 1–14. doi: 10.1038/s41598-017-08118-3.
- Carr, M. (2013) 'The water relations and irrigation requirements of olive (*Olea europaea* L.): A review', *Experimental Agriculture*, 49(04). doi: 10.1017/S0014479713000276.
- Conceição, N. *et al.* (2017) 'Evapotranspiration measured in a traditional rainfed and an irrigated intensive olive orchard during a year of hydrological drought.', *Acta Horticulturae*, (1150), pp. 281–288.
- Conceição, N. *et al.* (2017) 'Three years of monitoring evapotranspiration components and crop and stress coefficients in a deficit irrigated intensive olive orchard', *Agricultural Water Management*, 191, pp. 138–152. doi: 10.1016/j.agwat.2017.05.011.
- Conceição, N. and Ferreira, M. I. (2009) 'Combination of sap flow and eddy covariance techniques to obtain long term transpiration in a pear orchard', *Acta Horticulturae*, 846, pp. 53–60. doi: 10.17660/ActaHortic.2009.846.3.
- Corell, M. *et al.* (2014) 'Using band dendrometers in irrigation scheduling. Influence of the location inside the tree and comparison with point dendrometer', *Agricultural Water Management*. Elsevier B.V., 142, pp. 29–37. doi: 10.1016/j.agwat.2014.04.005.
- Cuevas, M. V. *et al.* (2010) 'Assessment of trunk diameter variation derived indices as water stress indicators in mature olive trees', *Agricultural Water Management*, 97(9), pp. 1293–1302. doi: 10.1016/j.agwat.2010.03.011.

- Doorenbos, J. and Pruitt, W. O. (1977) 'Crop water requirements.', *Irrigation and Drainage Paper No.24*.
- Durugkar, I. P. et al. (2016) 'Deep Reinforcement Learning With Macro-Actions', pp. 2094–2100. Available at: <http://arxiv.org/abs/1606.04615>.
- Dutta, S. (2018) *Reinforcement Learning with TensorFlow*.
- Faris, J. D. (2013) 'Wheat Domestication: Key to Agricultural Revolutions Past and Future', *Genomics of Plant Genetic Resources*, pp. 439–464.
- Fernández, J. E. and Cuevas, M. V. (2010) 'Irrigation scheduling from stem diameter variations: A review', *Agricultural and Forest Meteorology*, 150(2), pp. 135–151. doi: 10.1016/j.agrformet.2009.11.006.
- Ferreira, M. et al. (2008) 'Evapotranspiration estimates and water stress indicators for irrigation scheduling in woody plants.', *Cap. 3 em Agricultural Water Management Research Trends*, p. 129\_170.
- Ferreira, M. et al. (2011) 'Transpiration and water stress effects on water use, in relation to estimations from NDVI: application in a vineyard in SE Portugal.', *Options Méditerranéennes: Série B*, (67), pp. 203–208.
- Ferreira, M. (2017) 'Stress Coefficients for Soil Water Balance Combined with Water Stress Indicators for Irrigation Scheduling of Woody Crops', *Horticulturae*, 3(2), p. 38. doi: 10.3390/horticulturae3020038.
- Ferreira, M., Green, S. and Conceição, N. (2018) 'Assessing hydraulic redistribution with the compensated average gradient heat-pulse method on rain-fed olive trees', *Plant Soil*.
- Ferreira, M. I. et al. (1996) 'Evapotranspiration and crop-water relations in a peach orchard', *Evapotranspiration and irrigation scheduling*, pp. 61–68.
- Ferreira, M. I. et al. (2012) 'Análise de indicadores de desconforto hídrico durante ciclos de stress num olival intensivo no Alentejo', *VI Actas do Simpósio Nacional de Olivicultura*, 21, pp. 207–2015.
- Geerts, S. and Raes, D. (2009) 'Deficit irrigation as an on-farm strategy to maximize crop water productivity in dry areas', *Agricultural Water Management*, 96(9), pp. 1275–1284. doi: 10.1016/j.agwat.2009.04.009.
- Géron, A. (2017) *Hands On Machine Learning With Scikit-Learn And Tensorflow*. O'REILLY MEDIA, INC, USA.
- Gerten, D. et al. (2011) 'Global Water Availability and Requirements for Future Food Production', *Journal of Hydrometeorology*, 12(5), pp. 885–899. doi: 10.1175/2011JHM1328.1.
- Ginestar, C. and Castel, J. R. (1998) 'Use of stem dendrometers as indicators of water stress in drip-irrigated citrus trees', *Acta Horticulturae*, pp. 209–219.
- Giourga, C. and Loumou, A. (2003) 'Olive groves: "The life and identity of the Mediterranean"', *Agriculture and Human Values*, 20(1), pp. 87–95.
- Girón, I. F. et al. (2016) 'Limitations and usefulness of maximum daily shrinkage (MDS) and trunk growth rate (TGR) indicators in the irrigation scheduling of table olive trees', *Agricultural Water Management*. Elsevier B.V., 164(2010), pp. 38–45. doi: 10.1016/j.agwat.2015.09.014.
- Goldhamer, D. A. and Fereres, E. (2001) 'Irrigation scheduling protocols using continuously recorded trunk diameter measurements', *Irrigation Science*, 20(3), pp. 115–125. doi: 10.1007/s002710000034.
- Gonzalez-Sanchez, a (2014) 'Predictive ability of machine learning methods for massive crop yield prediction', *Spanish Journal of Agricultural Research*, 12(2), pp. 313–328. doi: 10.5424/sjar/2014122-4439.
- Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. MIT Press.
- Hausknecht, M. and Stone, P. (2015) 'Deep Recurrent Q-Learning for Partially Observable MDPs', pp. 29–37. doi: 10.1.1.696.1421.
- Hopfield, J. J. (1988) 'Artificial neural networks', *IEEE Circuits and Devices Magazine*, 4(5), pp. 3–10.
- Horrigan, L., Lawrence, R. S. and Walker, P. (2002) 'How sustainable agriculture can address the environmental and human health harms of industrial agriculture', *Environmental Health Perspectives*, 110(5), pp. 445–456. doi: 10.1289/ehp.02110445.



*How Neural Networks Work* (2017).

*International Olive Council* (2018). Available at: <http://www.internationaloliveoil.org> (Accessed: 10 September 2018).

Jensen, M. E., Burmann, R. D. and Allen, R. G. (2016) *Evaporation and irrigation water requirements, ASCE manual and reports on engineering practice*. doi: 10.1061/9780784414057.

Kamilaris, A., Kartakoullis, A. and Prenafeta-Boldú, F. X. (2017) 'A review on the practice of big data analysis in agriculture', *Computers and Electronics in Agriculture*. Elsevier, 143(January), pp. 23–37. doi: 10.1016/j.compag.2017.09.037.

Karandish, F. and Šimůnek, J. (2016) 'A comparison of numerical and machine-learning modeling of soil water content with limited input data', *Journal of Hydrology*, 543, pp. 892–909. doi: 10.1016/j.jhydrol.2016.11.007.

Kriesel, D. (2014) *A Brief Introduction to Neural Networks*.

Kullberg, E. G., DeJonge, K. C. and Chávez, J. L. (2017) 'Evaluation of thermal remote sensing indices to estimate crop evapotranspiration coefficients', *Agricultural Water Management*. Elsevier B.V., 179, pp. 64–73. doi: 10.1016/j.agwat.2016.07.007.

Liu, D., Yu, Z. and Lü, H. (2010) 'Data assimilation using support vector machines and ensemble Kalman filter for multi-layer soil moisture prediction', *Water Science and Engineering*. Hohai University. Production and hosting by Elsevier B.V., 3(4), pp. 361–377. doi: 10.3882/j.issn.1674-2370.2010.04.001.

Lorite, I. J. *et al.* (2007) 'Assessing deficit irrigation strategies at the level of an irrigation district', *Agricultural Water Management*, 91(1–3), pp. 51–60. doi: 10.1016/j.agwat.2007.04.005.

McCarthy, A. C., Hancock, N. H. and Raine, S. R. (2014) 'Development and simulation of sensor-based irrigation control strategies for cotton using the VARlwise simulation framework', *Computers and Electronics in Agriculture*. Elsevier B.V., 101, pp. 148–162. doi: 10.1016/j.compag.2013.12.014.

Miller, G. A. *et al.* (2014) 'Field evaluation and performance of capacitance probes for automated drip irrigation of watermelons', *Agricultural Water Management*. Elsevier B.V., 131, pp. 124–134. doi: 10.1016/j.agwat.2013.09.012.

Mitchell, T. (1997) *Machine Learning*. McGraw Hill.

Mnih, V., Graves, A. and Silver, D. (2016) 'Asynchronous Methods for Deep Reinforcement Learning', *International Conference on Machine Learning*, 48.

Molden, D. *et al.* (2010) 'Improving agricultural water productivity: Between optimism and caution', *Agricultural Water Management*, 97(4), pp. 528–535. doi: 10.1016/j.agwat.2009.03.023.

Nadezhkina, N. *et al.* (2010) 'Trees never rest: the multiple facets of hydraulic redistribution', *Ecohydrology*.

Nadezhkina, N. *et al.* (2015) 'Water uptake and hydraulic redistribution under a seasonal climate: Long-term study in a rainfed olive orchard', *Ecohydrology*, 8(3), pp. 387–397. doi: 10.1002/eco.1545.

Naor, A. and Cohen, S. (2003) 'Sensitivity and variability of maximum trunk shrinkage, midday stem water potential, and transpiration rate in response to withholding irrigation from field-grown apple trees', *HortScience*, 38(4), pp. 547–551.

Nogueira, A. M. *et al.* (2012) 'Pegada hídrica de um olival super-intensivo em condições mediterrânicas, utilizando medições de campo e detecção remota.', 21, p. 8. Available at: [dspace.uevora.pt/rdpc/handle/10177/11111/1/nogueira\\_et\\_al\\_paper\\_VI\\_simpOliv.pdf](https://dspace.uevora.pt/rdpc/handle/10177/11111/1/nogueira_et_al_paper_VI_simpOliv.pdf).

*O que é o neurónio?* (no date).

Oliveira, C. (2016) 'Aulas de Fruticultura ISA 2016'.

Paço, T. A. *et al.* (2011) 'Vineyard evapotranspiration measurement in non-flat terrain conditions by the Eddy covariance technique -Douro Wine Region, Portugal', *Acta Horticulturae*, 922, pp. 295–302. doi: 10.17660/ActaHortic.2011.922.38.

Palese, A. M. *et al.* (2010) 'Effects of water deficit on the vegetative response, yield and oil quality of olive trees (*Olea europaea* L., cv Coratina) grown under intensive cultivation', *Scientia Horticulturae*. Elsevier B.V., 125(3), pp. 222–229. doi: 10.1016/j.scienta.2010.03.025.

Passioura, J. B. (1988) 'Response to Dr P.J. Kramer's article, "Changing concepts regarding plant water relations"', *Plant Cell Environ*, 11, pp. 569–571.

- Pauwels, V. R. N. and Samson, R. (2006) 'Comparison of different methods to measure and model actual evapotranspiration rates for a wet sloping grassland', *Agricultural Water Management*, 82(1–2), pp. 1–24. doi: 10.1016/j.agwat.2005.06.001.
- Pereira, L. S. et al. (2002) 'Irrigation management under water scarcity', *Agricultural Water Management*, 57(3), pp. 175–206. doi: 10.1016/S0378-3774(02)00075-6.
- Pereira, L. S. and Alves, I. (2013) *Crop Water Requirements, Reference Module in Earth Systems and Environmental Sciences*. Elsevier Inc. doi: 10.1016/B978-0-12-409548-9.05129-0.
- Perrier, A. (1985) 'Updated evapotranspiration and crop water requirement definitions', *Crop Water Requirements*, pp. 885–887.
- Pokharna, H. (2016) *The Introduction to Neural Networks we all need*.
- Qadir, M. et al. (2003) 'Agricultural water management in water-starved countries: Challenges and opportunities', *Agricultural Water Management*, 62(3), pp. 165–185. doi: 10.1016/S0378-3774(03)00146-X.
- Rijsberman, F. R. (2006) 'Water scarcity: Fact or fiction?', *Agricultural Water Management*, 80(1–3 SPEC. ISS.), pp. 5–22. doi: 10.1016/j.agwat.2005.07.001.
- Robinson, D. A., Gardner, C. M. K. and Cooper, J. D. (1999) 'Measurement of relative permittivity in sandy soils using TDR, capacitance and theta probes: Comparison, including the effects of bulk soil electrical conductivity', *Journal of Hydrology*, 223(3–4), pp. 198–211. doi: 10.1016/S0022-1694(99)00121-3.
- Sadras, V. O. and Milroy, S. P. (1996) 'Soil-water threshold for the responses of leaf expansion and gas exchange: a review.', *Field Crop research*, 47, pp. 253–266.
- Saito, S., Wenzhuo, Y. and Shanmugamani, R. (2018) *Python Reinforcement Learning Projects*. Packt.
- Santos, F. L. et al. (2007) 'Water use and response of a dry-farmed olive orchard recently converted to irrigation', *Biosystems Engineering*, 98(1), pp. 102–114. doi: 10.1016/j.biosystemseng.2007.03.027.
- Santos, F. L. et al. (2013) 'Efeitos da rega e do regime hídrico em olival super intensivo no Alentejo Effects of irrigation and water regime in super-intensive olive orchard in Alentejo', 36(2007), pp. 206–219.
- Sanz-Cortés, F. et al. (2002) 'Phenological growth stages of olive trees (*Olea europaea*)', *Annals of Applied Biology*, 140(2), pp. 151–157. doi: 10.1111/j.1744-7348.2002.tb00167.x.
- Shalev-Shwartz, S. and Ben-David, S. (2014) *Understanding Machine Learning*. Cambridge University Press.
- Shukla, A. et al. (2014) 'Soil Moisture Estimation using Gravimetric Technique and FDR Probe Technique: A Comparative Analysis', *ISSN ISSN (CD-ROM American International Journal of Research in Formal, Applied & Natural Sciences AIJRFANS)*, (January), pp. 2328–3777. Available at: <http://www.iasir.net>.
- Silva, R. (2009) *Desenvolvimento de um Sistema Inteligente de Determinação das Necessidades Hídricas para Culturas Lenhosas Anisotrópicas*. Technical University of Lisbon. Instituto Superior de Agronomia.
- Silva, R. and Ferreira, M. (2014) 'A new microdendrometric index to quantify water status in kiwifruit.', *Acta Horticulturae*, (1038), pp. 43–50.
- Silver, D. et al. (2017) 'Mastering the game of go without human knowledge', *Nature*, (550(7676)), p. 354.
- Solheim II, W. G. (1972) 'AN EARLIER AGRICULTURAL REVOLUTION', *Scientific American*, 226(4), pp. 34–41.
- Sonka, S. (2014) 'Big data and the ag sector: More than lots of numbers', *International Food and Agribusiness Management Review*, 17(1), pp. 1–20.
- Sutton, R. S. and Barto, A. G. (1998) *Reinforcement Learning: An Introduction*. MIT Press.
- Synced (2017) *Big Picture Machine Learning: Classifying Text with Neural Networks and TensorFlow*. Available at: <https://medium.com/syncedreview/big-picture-machine-learning-classifying-text-with-neural-networks-and-tensorflow-da3358625601>.
- Taylor, P. (2000) 'Water Use by the Olive Tree', *Journal of Crop Production*, (908159515), pp. 37–41. doi: 10.1300/J144v02n02.

- Tilman, D. *et al.* (2002) 'Agricultural sustainability and intensive production practices', *Nature*, 418(6898), pp. 671–677. doi: 10.1038/nature01014.
- Topp, G. C. and Davis, J. L. (1985) 'Measurement of soil water content using time-domain reflectometry (TDR): A Field Evaluation', *Soil Science Society of America Journal*, 49, pp. 19–24.
- Turner, N. (1986) 'Adaptation to Water Deficits: A Changing Perspective.', *J. Plant Physiol. Aust.*, 13, pp. 175–190.
- Veihmeyer, F. J. and Hendrickson, A. H. (1949) 'METHODS OF MEASURING FIELD CAPACITY AND PERMANENT WILTING PERCENTAGE OF SOILS', *Soil Science*, 68(1), pp. 75–94.
- Vera, J. *et al.* (2009) 'Soil water balance trial involving capacitance and neutron probe measurements', *Agricultural Water Management*, 96(6), pp. 905–911. doi: 10.1016/j.agwat.2008.11.010.
- Veraa, J. *et al.* (2017) 'Precise sustainable irrigation: A review of soil-plant-atmosphere Monitoring', *Acta Horticulturae*, 1150, pp. 195–200. doi: 10.17660/ActaHortic.2017.1150.28.
- Vieira, M. J. P. (2003) 'Gestão da Água em Portugal. Os Desafios do Plano Nacional da Água', *Engenharia Civil*, 16, pp. 5–13. doi: 10.2307/3146872.
- Vienken, T. *et al.* (2013) 'Field comparison of selected methods for vertical soil water content profiling', *Journal of Hydrology*. Elsevier B.V., 501, pp. 205–212. doi: 10.1016/j.jhydrol.2013.08.004.
- Wallace, J. S. (2000) 'Increasing agricultural water use efficiency to meet future food production', *Agriculture, Ecosystems and Environment*, 82(1–3), pp. 105–119. doi: 10.1016/S0167-8809(00)00220-6.
- Wang, Z. *et al.* (2015) 'Dueling Network Architectures for Deep Reinforcement Learning', (9). doi: 10.1109/MCOM.2016.7378425.
- Watkins, C. J. C. H. and Dayan, P. (1992) 'Q-learning', *Machine Learning*, 8(3–4), pp. 279–292. doi: 10.1007/BF00992698.
- Zaccone, G. and Karim, R. (2018) *Deep Learning with TensorFlow*. Packt.
- Zweifel, R., Item, H. and Häslar, R. (2001) 'Link between diurnal stem radius changes and tree water relations', *Tree Physiology*, 21(12–13), pp. 869–877. doi: 10.1093/treephys/21.12-13.869.
- Zweifel, R., Zimmermann, L. and Newbery, D. M. (2005) 'Modeling tree water deficit from microclimate: an approach to quantifying drought stress RID C-2852-2009 RID B-4508-2008', *Tree Physiology*, 25(2), pp. 147–156.